DOI: 10.1111/phpr.12742

ORIGINAL ARTICLE

Philosophy and Phenomenological Research

Externalism and exploitability

Nilanjan Das

University College London

Correspondence
Nilanjan Das
Email: nilanjan.das@ucl.ac.uk

Abstract

According to Bayesian orthodoxy, an agent should update—or at least should plan to update—her credences by conditionalization. Some have defended this claim by means of a *diachronic Dutch book argument*. They say: an agent who does not plan to update her credences by conditionalization is vulnerable (by her own lights) to a diachronic Dutch book, i.e., a sequence of bets which, when accepted, guarantee loss of utility. Here, I show that this argument is in tension with *evidence externalism*, i.e., the view that an agent's evidence can entail non-trivial propositions about the external world. I argue that this tension casts doubt on the idea that diachronic Dutch books can be used to justify or vindicate updating plans.

Some think that epistemic rationality consists in the instrumentally rational pursuit of some kind of value. Call this view *instrumentalism*. Depending on what sort of value we take to be relevant to epistemic rationality, instrumentalism can come in different varieties. Some take the relevant kind of value to be distinctively epistemic, e.g., truth or accuracy. *Accuracy-first epistemology* is an instance of this approach. Accuracy-firsters think that norms of epistemic rationality can be justified (or vindicated) on the grounds that violating them makes us vulnerable to sure or expected loss of accuracy.[1] Others think that the relevant kind of value need not be distinctively epistemic; it could be the kind of pragmatic value that instrumentally rational agents promote when they perform actions. *Dutch book arguments* are motivated by this thought. Defenders of such arguments think that norms of epistemic rationality can be justified by appealing to the fact that violating them makes us vulnerable to sure loss

[1]For instances of this approach, see Joyce (1998, 2009), Greaves and Wallace (2006), Leitgeb and Pettigrew (2010a, 2010b), Pettigrew (2016), and Briggs and Pettigrew (2020).

of utility.[2] The aim of this paper is to offer an argument, which, together with other arguments in the existing literature, reveals a limitation of instrumentalism: instrumentalism makes undesirable predictions about which updating plans are epistemically rational for us to adopt.

My starting point will be an updating plan that is commonly endorsed by Bayesians. According to Bayesian orthodoxy, an agent should update, or should plan to update, her credences by *conditionalization*. If an agent updates by conditionalization, then, in situations of evidence-gathering where she gains new evidence without losing any evidence, her posterior credence in any proposition matches her *prior conditional credence* in that proposition given her new evidence. We shall focus on:

> *Plan Conditionalization*. If an agent is rationally certain that, in a scenario of evidence-gathering, she might gain new evidence but won't lose any evidence, then she is required by epistemic rationality to plan to update her credences by conditionalization in that scenario.

Some defend *Plan Conditionalization* by appealing to a *diachronic Dutch book argument* (DDBA).[3] A diachronic Dutch book is a system of bets which are offered at multiple points of time and which, when accepted, result in sure utility loss. According to the DDBA for *Plan Conditionalization*, if an agent does not plan by update by conditionalization, she is vulnerable (by her own lights) to a diachronic Dutch book.

The defenders of this DDBA presuppose that foreseeable exploitability indicates irrationality, i.e., that, if an agent is rationally certain that complying with a certain updating plan will make her vulnerable to sure loss of utility, then it is epistemically irrational for her to adopt that plan. Here, I show that this assumption is in tension with the combination of *Plan Conditionalization* and an attractive conception of evidence:

> *Evidence Externalism*. An agent's evidence is a proposition or a set of propositions that may entail non-trivial propositions about the external world.

The argument is this. If *Evidence Externalism* is true, we have to countenance cases where an agent's evidence before an inquiry leaves open either the possibility that her future evidence will entail falsehoods, or that she will not have perfect access to her future evidence. If the agent updates by conditionalization in such scenarios, she will be subject to a fairly strong diachronic Dutch book. So, if foreseeable exploitability indicates irrationality, then *Plan Conditionalization* is false.

This argument exposes a problem with instrumentalism. DDBAs and accuracy-based arguments for *Plan Conditionalization* show that an agent is rationally required to adopt the plan of updating by conditionalization under certain conditions, i.e., when her evidence before an inquiry entails that her future evidence will only entail truths and that she will have perfect access to her future evidence.[4] A number of writers have noticed that

[2]For arguments of this kind, see Ramsey (1926), de Finetti (1937), Lehman (1955) and Kemeny (1955), Teller (1973), Skyrms (1987) and Lewis (1999).

[3]Lewis (1999) was the first proponent of this argument, though he wanted the argument to support the simpler claim that an agent is required by epistemic rationality to update by conditionalization. Teller (1973) reported Lewis' argument, and van Fraassen (1984) explained why Lewis' argument supports a norm that requires the agent to commit to an updating plan rather than a norm that applies directly to updating. Apart from these, there are other arguments for *Plan Conditionalization*. Williams (1980) uses the Principle of Minimum Information to defend *Plan Conditionalization*. Van Fraassen (1999, 1989) appeals to certain symmetry considerations to argue for it. More recently, Oddie (1997), Greaves and Wallace (2006), Easwaran (2013), and Briggs and Pettigrew (2020) have offered accuracy-based arguments for it.

[4]For accuracy-based arguments for *Plan Conditionalization*, Oddie (1997) and Greaves and Wallace (2006). Briggs and Pettigrew (2020) turn this into an accuracy-dominance argument for the diachronic norm of updating by conditionalization.

DDBAs and accuracy-based arguments do not support *Plan Conditionalization* when some of these conditions are relaxed.[5] What I show is that the tension amongst *Plan Conditionalization*, *Evidence Externalism* and the assumption that foreseeable exploitability indicates irrationality cannot easily be resolved by rejecting either *Evidence Externalism* or *Plan Conditionalization*. This suggests that we should reject the assumption that being foreseeably exploitable is always a mark of irrationality. As Schoenfield's (2017) work suggests, the same line of reasoning can be replicated for the accuracy-first approach. This yields some evidence that we cannot justify (or vindicate) updating plans by relying on instrumentalist approaches to epistemic rationality.

Here is how I shall proceed. After introducing some useful technical notions (§1), I will show that there are two constraints—*transparency* and *reflectiveness*—that are individually necessary and jointly sufficient for an updating plan to be immune from a diachronic Dutch book (§2). When an agent's evidence entails that her future evidence will only entail truths and that she will have perfect access to her future evidence, the plan of updating by conditionalization is the only plan that is both transparent and reflective by her own lights. In such circumstances, therefore, an agent can escape diachronic Dutch books by her own lights only by adopting that plan (§3). However, I will argue that the defender of *Evidence Externalism* must countenance cases where the plan of updating by conditionalization foreseeably fails to be either transparent or reflective. In such cases, if the agent plans to update by conditionalization, she will vulnerable by her own lights to diachronic Dutch books (§4). In the rest of the paper, I will explain why this is bad news for defenders of DDBAs (§5-7).[6]

1 | HOUSEKEEPING

In this section, I will introduce a few useful technical notions.

1.1 | Inquiries, priors, and plans

Let an *inquiry* be any evidence-gathering act where the agent's evidence before the act entails that she will engage in that act but will not lose any evidence while doing so. We will represent any inquiry with a structure $\langle W, E \rangle$. W is a finite set of worlds that captures the agent's evidence before the inquiry.[7] E is an *evidence function* that maps each world in W to a proposition (i.e., a set of worlds in W) which is the strongest piece of posterior evidence that the agent receives in that world.

[5]For this observation in relation to DDBAs, see Hild (1998a, 1998b), Bronfman (2014) and Gallow (2019). For this observation in relation to accuracy-based arguments, see Schoenfield (2017), Bronfman (2014) and Das (2019), and Gallow (forthcoming).

[6]I want to be clear about the scope of my argument. Some have expressed scepticism about the probative value of Dutch book arguments. For concerns of this sort about *synchronic* Dutch book arguments for Probabilism (i.e., the norm of maintaining probabilistically coherent credences), see Christensen (1991) and Pettigrew (2020). For responses, see Skyrms (1987), Christensen (1996), Mahtani (2015) and Pettigrew (2020). It is controversial whether similar responses are available for diachronic Dutch books; for discussion, see Briggs (2009) and Mahtani (2012). Here, I will only explore whether foreseeable *dynamic* exploitability—which is exposed by *diachronic* Dutch books—can be used to support *Plan Conditionalization*. So, my arguments leave open the possibility that some Dutch book arguments (e.g., synchronic ones) can support epistemic norms like Probabilism.

[7]The assumption that W is finite might seem a little artificial: it is natural to think that the sets of all possible worlds that are compatible with an agent's evidence ought to be uncountably infinite. However, I take this to be a harmless idealization; with some additional constraints, all the results proved in this paper can be generalized to inquiries that involve infinite sets of possible worlds.

To fill out the picture a bit more, we will need the notion of a rational *prior credence function*. I will assume that a rational prior credence function is a *regular* probability function that maps any proposition to a real value.[8] This reflects the degree of belief or credence that the agent should have in that proposition before her inquiry.

This brings us to the third component of our formal framework: *updating plans*. An updating plan is a plan that tells an agent how to respond to the evidence that she receives. For any inquiry $\langle W, E \rangle$, it is natural to think of such a plan as a function U from worlds in to credence functions, such that, for any two worlds w, w^* in W, if the agent's evidence in the two worlds is the same, then U recommends the same credence function in them. In other words, if $E(w) = E(w^*)$, $U(w) = U(w^*)$.[9] To see how the notion works, consider *conditionalizing plans*, i.e., updating plans that require the agent to update by conditionalization. Relative to an inquiry $\langle W, E \rangle$ and any rational prior credence function p, an updating plan U is a conditionalizing plan just in case, for any world w in W, $U(w) = p(\,.\,|\,E(w)\,)$ (provided $p(E(w)) > 0$). It is easy to check that conditionalizing plans are updating plans in the sense specified above: whenever an agent's evidence is the same in any two worlds in W, a conditionalizing plan will recommend the same credence function in those two worlds.

1.2 | Diachronic Dutch books

Finally, I can introduce the last component of our formal framework: *diachronic Dutch books*.

Let a *bet* be a gamble that yields a certain (positive or non-positive) payoff if a certain proposition is true, but nothing otherwise. The amount of utility that it pays is the *stake* of the bet. A *book* is just a collection of bets $G_1, G_2, \ldots, G_n$ on a set of propositions $\{P_1, P_2, \ldots, P_n\}$. Thus, we may think of a book B as a function from propositions to stakes: if a book B is a collection of bets $G_1, G_2, \ldots, G_n$ on a set of propositions $\{P_1, P_2, \ldots, P_n\}$, then B maps each P_i to the stake of the corresponding bet G_i on that proposition.

We can calculate the payoff of a book B in any world w just by summing up the stakes of those bets in B which are offered on propositions that are true in w. Suppose B is a collection of three bets G_1, G_2 and G_3 with stakes of s_1, s_2, and s_3 on three propositions P_1, P_2, and P_3 respectively. If only P_1 and $P_2$are true in w, then the payoff of the book will just be the sum of s_1 and s_2. Let's put this a bit more precisely. For any inquiry $\langle W, E \rangle$, let v be a function such that for any $w \in W$ and any proposition $P \subseteq W$, $v(P, w) = 1$ if $w \in P$ and $v(P, w) = 0$ if $w \notin P$. Then, if B is a book defined on a set of propositions $\{P_1, P_2, \ldots, P_n\}$, then, for any $w \in W$, the payoff of B in w is $\texttt{payoff}\ \ (B, w) = \sum_{i=1}^{n} B(P_i)\, v(P_i, w)$.

[8]For any inquiry $\langle W, E \rangle$, a regular credence function p defined on the subsets of W is one that assigns non-zero credence to any singleton set containing a world in W. For defences of regularity as a general constraint of epistemic rationality irrespective of the size of the possibility space, see Lewis (1980), Skyrms (1980), and McGee (1994). For arguments against this general constraint, see Williamson (2007) and Easwaran (2014). For my purposes, I only need a weaker requirement: namely, if an agent distributes her credences over a finite set of possibilities, she shouldn't assign non-zero credence to a possibility when her evidence does not rule it out.

[9]This ensures that an updating plan does not require the agent to adopt different credence functions even when her evidence does not vary. For discussion, see Greaves and Wallace (2006) and Schoenfield (2017).

Next, I will assume that the maximum price that an agent can rationally pay for a book is the weighted average of the stakes of the bets in that book, where the weight for each bet is the credence that the agent assigns to the corresponding proposition.[10] So, if B is a book defined on a set of propositions $\{P_1, P_2, \ldots, P_n\}$, then the maximum price that an agent rationally pay for the book B relative to a credence function c is $\texttt{price}\,(B, c) = \sum_{P_i} c(P_i)B(P_i)$.[11]

Using the notions of payoff and maximum price, we can define a notion of *minimum profit*. The minimum profit that an agent can rationally make from a book in a world is just the payoff of that book in that world *minus* the maximum price that she can rationally pay for that book relative to her credence function. So, if B is a book defined on a set of propositions $\{P_1, P_2, \ldots, P_n\}$, the minimum profit that an agent can rationally make from B at a world w given her credence function c is $\texttt{profit}\,(B,c,w) = \texttt{payoff}\,(B,w) - \texttt{price}\,(B,c)$.

Using these three notions, we can finally state what a diachronic Dutch book is. Suppose the strongest pieces of evidence that an agent could receive as a result of an inquiry are $E_1, E_2, \ldots, E_k$. A diachronic Dutch book for an updating plan U is a series of books $B, B_1, \ldots, B_k$ where (i) B is sold to an agent updating according to U at maximum price before she gathers evidence, (ii) each B_i is sold to that agent at maximum price when the strongest piece of future evidence she receives is E_i, and (iii) as a result, the agent incurs guaranteed loss of utility. More precisely,

> *Diachronic Dutch Books*. For any inquiry $\langle W, E\rangle$, let p be the relevant agent's rational prior credence function, and let $\mathscr{E} = \{E(w) : w \in W\} = \{E_1, E_2, \ldots, E_k\}$ be the set of the strongest pieces of evidence that the agent could gain as a result of her inquiry. An updating plan U defined on that inquiry is subject to a diachronic Dutch book iff there exist books $B, B_1, \ldots, B_k$, such that, for any $w \in W$, if $E(w) = E_i$, $\texttt{profit}\,(B, p, w) + \texttt{profit}\,\left(B_i, U(w), w\right) < 0$.

For our purposes, it will be useful to divide diachronic Dutch books into two kinds: *invariant* and *variant*. In some cases where an agent's updating plan is subject to a diachronic Dutch book, she is offered different books after she has gathered evidence depending on what her posterior evidence is. Let an *invariant* diachronic Dutch book be a diachronic Dutch book where there is no such variation. Thus, an invariant diachronic Dutch book for an updating plan U simply consists of two books B_1 and B_2 where (i) B_1 is sold to an agent updating according to U at maximum price before she gathers evidence, (ii) B_2 is sold to her at maximum price after she has gathered evidence, and (iii) as a result, she incurs sure loss of utility. Let any diachronic Dutch book that is not invariant be *variant*.

In the next section, I will explore the conditions under which updating plans are subject to variant or invariant diachronic Dutch books.

[10] Implicitly, I am making two assumptions. First, or any proposition P, if an agent has credence r in P, she can rationally pay an amount no greater than $r.s$ for a bet that has a stake of s on P. Second, for any collection of bets $G_1, G_2, \ldots, G_n$, if the maximum price that an agent can rationally pay for each bet G_i is p_i, then the maximum price that the agent rationally pay for the entire collection is just the sum of the p_i's. For discussion of the first principle, see Hedden (2013), Wrónski and Godziszewski (2017) and Pettigrew (forthcoming). For discussion of the second principle, see Buchak (2013, Chapter 7).

[11] When the credence function c is a probability function, this will just be the expected payoff of the book. In those cases, therefore, $\texttt{price}(B, c) = \sum_{w \in W} c(w)\texttt{payoff}(B, w)$. However, when the agent is probabilistically incoherent, the maximum price of the book needn't be equivalent to its expected payoff.

2 | TRANSPARENCY AND REFLECTIVENESS

There are two conditions that are jointly necessary and sufficient for an updating plan to be immune from variant or invariant diachronic Dutch books: *transparency* and *reflectiveness*.

Start with *transparency*. A *transparent* plan is a plan such that, if an agent complies with it, she becomes certain about what evidence she has gained. For any inquiry $\langle W, E \rangle$, let $[\mathbf{E} = P]$ be the proposition that the strongest evidence that the agent has gained is P. So, it is just the set of worlds w where $E(w)$ is P. If an agent complies with a transparent plan, and the strongest posterior evidence that she has gained is P, then her posterior credence function will assign credence of 1 to $[\mathbf{E} = P]$. More precisely,

> *Transparency*. For any inquiry $\langle W, E \rangle$, an updating plan U is transparent iff, for any world $w \in W$, if $U(w) = c$, then $c([\mathbf{E} = E(w)]) = 1$.

Focus now on *reflectiveness*. This is best understood as an anti-bias constraint. A biased inquiry is an inquiry which, by the agent's own lights, is guaranteed to increase or decrease her credence in a proposition. Suppose an agent starts out with a prior credence of 0.5 in a proposition P, but adopts a plan which, by lights of her prior evidence, recommends the possible future credences of 0.6, 0.7 and 0.8 in P. In that scenario, the agent has adopted a plan which, by her own lights, is guaranteed to increase her credence in P. Thus, her plan allows her to bias her inquiry in favour of P. Updating plans that license such biased inquiry violate a really weak Reflection Principle. According to this principle, an agent's prior credence in any proposition P should be a weighted average of her planned future credences in P (where the weights are probabilities). We'll call plans that satisfy this constraint *reflective*.

> *Reflectiveness*. For any inquiry $\langle W, E \rangle$, let p be the relevant agent's rational prior credence function, and let $\mathscr{E} = \{E(w) : w \in W\} = \{E_1, E_2, \ldots, E_k\}$ be the set of the strongest pieces of evidence that the agent could gain as a result of her inquiry. Let U be an updating plan defined on that inquiry, such that U_i is the credence function recommended by U relative to E_i. Then, U is reflective (relative to p) iff there exists a probability distribution p^* such that, for any proposition P, [12]

$$p(P) = \sum_{i=1}^{k} U_i(P) p^*([\mathbf{E} = E_i]).$$

To see how the notion of reflectiveness works, consider the case where the agent's prior credence in a proposition is 0.5, but her planned future credences in that very proposition are 0.6, 0.7 and 0.8. Since 0.5 cannot be expressed as a weighted average of 0.6, 0.7. and 0.8 (provided that the weights are probabilities), the agent's plan is not reflective.

[12] Skyrms (1987) calls *reflectiveness* the *interior condition*; for further discussion, see Pettigrew (2020, §8). The constraint of reflectiveness needs to be distinguished from two other Reflection principles. The first is similar to van Fraassen's (1984) Reflection Principle. This principle says that an agent's prior conditional credence in P given that her future evidence is E_i should just be her planned future credence in P relative to E_i, i.e., $p(P \mid [\mathbf{E} = E_i]) = U_i(P)$. The other generalized version of van Fraassen's principle is a bit more complicated. It says that the agent's prior credence in P should be her prior expectation of her planned future credences in P, i.e., $p(P) = \sum_{i=1}^{k} U_i(P) p([\mathbf{E} = E_i])$. The difference between the constraint of reflectiveness and these two principles lies in this. If a plan is reflective, then, for any P, the agent's prior credence in P is a weighted average of her planned credences in P, but it doesn't have to match her prior expectation of her planned credences in P. Both van Fraassen's Reflection Principle and the generalized version of it impose the stronger requirement.

Let an updating plan be *probabilistic* just in case it only recommends probability functions relative to different bodies of evidence. There is a natural class of probabilistic plans that are both transparent and reflective: the *meta-conditionalizing plans*. Relative to an inquiry $\langle W, E \rangle$ and a rational prior credence function p, an updating plan U is a *meta-conditionalizing plan* just in case, for any world w, $U(w) = p(. \mid [\mathbf{E} = E(w)])$. We can show:

Proposition 1 *Relative to an inquiry $\langle W, E \rangle$ and any rational prior credence function p, let U be a probabilistic updating plan defined on that inquiry. Then, U is both transparent and reflective relative to p iff U is a meta-conditionalizing plan relative to p.*[13]

The difference between a conditionalizing plan and a meta-conditionalizing plan is this. When the strongest evidence that an agent gains is E_i, the conditionalizing plan requires the agent to adopt as her posterior credence in any proposition P her prior conditional credence in P given E_i. In contrast, in that scenario, a meta-conditionalizing plan requires the agent to adopt as her posterior credence in P her prior conditional credence in P given the proposition that the strongest evidence she has gained is E_i. Therefore, in cases where E_i is not the same proposition as the proposition that the strongest evidence that the agent has gained is E_i, the recommendations of conditionalizing plans will come apart from the recommendations of meta-conditionalizing plans.

We are now in a position to prove a number of results about transparent and reflective plans. First, if a plan is not transparent, then it is subject to a *variant* diachronic Dutch book, i.e., a diachronic Dutch book where the bookie offers different books depending on the agent's posterior evidence. Second, if an updating plan is not reflective, then it is subject to an invariant diachronic Dutch book. Third, if a plan is both reflective and transparent, it is subject to no diachronic Dutch book whatsoever. Let's state these claims more carefully.

Proposition 2 *For any inquiry and any rational prior credence function , let be a probabilistic updating plan defined on that inquiry. Then, the following three claims are true.*

- *If U is not transparent, then U is subject to a variant diachronic Dutch book.*
- *If U is not reflective relative to p, then U is subject to an invariant diachronic Dutch book.*
- *If U is transparent and reflective relative to p, then U is not subject to a (variant or invariant) diachronic Dutch book.*

Some of these claims are parasitic on existing results. The first claim in *Proposition 2* can be proved by adapting the Dutch book argument that Lewis (1999) and Teller (1973) gave for conditionalization, while the third claim can be derived from Skyrms' (1987) converse Dutch book theorem for conditionalization.[14]

Together, *Propositions 1 and 2* imply:

[13] All proofs are in the appendix.

[14] However, there are some differences between the proofs that these authors give and the proofs that I give here (in the appendix). First, Lewis and Teller don't distinguish transparency and reflectiveness. As a result, they don't distinguish the different kinds of diachronic Dutch books that failures of transparency and failures of reflectiveness give rise to. My proof of *Proposition 2* does that. Second, even though my proof of the third claim makes use of the proof strategy underlying Skyrms' (1987) converse Dutch book theorem for conditionalization, I have tried to make clear how that strategy can be adapted for meta-conditionalizing plans.

Corollary 1 *For any inquiry $\langle W, E \rangle$ and any rational prior credence function p, let U be a probabilistic updating plan defined on that inquiry. Then, U is a meta-conditionalizing plan relative to p iff U is not subject to a diachronic Dutch book.*

The lesson: all and only meta-conditionalizing plans are able to escape diachronic Dutch books. A version of *Corollary 1* was proved by Hild (1998a).[15] An accuracy-based analogue of this corollary has been proved by Schoenfield (2017): namely, that all and only meta-conditionalizing plans maximize expected accuracy.

3 | THE DDBA FOR PLAN CONDITIONALIZATION

Building on *Corollary 1*, we can offer a DDBA for *Plan Conditionalization*. The argument requires two assumptions.

The first assumption is that, before any inquiry, an agent's evidence should entail that her future evidence will only entail truths, and that she will have perfect access to whatever future evidence she has. This assumption, in turn, can be broken down into three separate assumptions: *Factivity*, *Positive Introspection*, and *Negative Introspection*.

> *Factivity*. In any inquiry, by lights of the agent's prior evidence, if her posterior evidence entails a proposition P, then P is true.
>
> *Positive Introspection*. In any inquiry, by lights of the agent's prior evidence, if her posterior evidence entails a proposition P, then her posterior evidence entails that her posterior evidence entails P.
>
> *Negative Introspection*. In any inquiry, by lights of the agent's prior evidence, if her posterior evidence does not entail a proposition P, then her posterior evidence entails that her posterior evidence does not entail P.

Factivity, *Positive Introspection*, and *Negative Introspection* correspond to the following properties of inquiries: reflexivity, transitivity, and euclideanness. An inquiry $\langle W, E \rangle$ is *reflexive* just in case, for any w in W, w is in $E(w)$. An inquiry $\langle W, E \rangle$ is *transitive* just in case, for any worlds w_1, w_2, w_3 in W, if w_2 is in $E(w_1)$ and w_3 is in $E(w_2)$, then w_3 is in $E(w_1)$. Finally, an inquiry $\langle W, E \rangle$ is *euclidean* just in case, for any worlds w_1, w_2, w_3 in W, if w_2 is in $E(w_1)$ and w_3 is in $E(w_1)$, then w_3 is in $E(w_2)$. When an inquiry has all three of these properties, it is *partitional*: in such an inquiry, the evidence function E imposes a partition on W where, for any world w, each $E(w)$ is a cell containing all and only those worlds in which the strongest piece of future evidence the agent receives is $E(w)$.

If *Factivity* is true, then all inquiries will be reflexive. Analogously, if *Positive Introspection* is right, then all inquiries must be transitive. Finally, if *Negative Introspection* is right, then all inquiries will be euclidean. So, if all three constraints are true, all inquiries will be partitional. Our first assumption is:

[15]Hild's proof strategy is different from mine. He proves *Corollary 1* by showing that all and only meta-conditionalizing plans, or plans that conform to what he calls *auto-epistemic conditionalization*, satisfy van Fraassen's (1984) Reflection Principle, and all and only plans that conform to this principle are immune to diachronic Dutch books. However, he does not separately discuss two conditions that we discuss here: reflectiveness and transparency.

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

Let's proceed to our second assumption. Suppose an agent plans to update according to a plan which, by her own lights, is subject to a diachronic Dutch book. So, she can foresee that, if she complies with that plan, she will incur sure loss of utility. If foreseeable exploitability of this kind indicates epistemic irrationality, then the agent is not epistemically rational to adopt that plan. This yields:

Exploitability. For any inquiry $\langle W, E \rangle$, let U be an updating plan defined on that inquiry. It is rationally permissible for the agent to adopt U before her inquiry only if U is not subject to a diachronic Dutch book by her own lights.

We are now in a position to offer our DDBA for *Plan Conditionalization*.

The argument depends on two facts. First, for any inquiry, all and only meta-conditionalizing plans are immune to diachronic Dutch books. This follows from *Corollary 1*. Second, if *Partitional Evidence* is true, then, in any inquiry where an agent does not lose evidence but could gain evidence, all and only conditionalizing plans are meta-conditionalizing plans (by her lights). We establish this, by proving:

Proposition 3 *Relative to an inquiry* $\langle W, E \rangle$, *let* $\mathscr{E} = \{E(w) : w \in W\}$ *be the set of all the strongest pieces of evidence that an agent could gain.* $\langle W, E \rangle$ *is partitional iff, for any* $P \in \mathscr{E}$, $P = [\mathbf{E} = P]$.

According to *Proposition 3*, in any partitional inquiry, conditionalizing on an evidence-proposition P is the same as conditionalizing on the fact that P is the strongest evidence one has received. So, whenever an inquiry is partitional, then, relative to it, a plan is conditionalizing if and only if it is meta-conditionalizing.

These two facts yield the conclusion that, if *Partitional Evidence* is true, then, in any inquiry where the agent does not lose evidence but could gain evidence, all and only conditionalizing plans are immune from diachronic Dutchbooks. Given our first assumption, i.e., *Partitional Evidence*, we can now conclude that in any scenario where an agent is certain that she will gain evidence without losing any evidence, all and only conditionalizing plans are immune from diachronic Dutchbooks. By *Exploitability*, then, it follows that an agent is rationally permitted only to adopt conditionalizing plans. This is the DDBA for *Plan Conditionalization*.

4 | EXTERNALISM, EXPLOITABILITY, AND CONDITIONALIZATION

This argument depends crucially on *Partitional Evidence*: if *Partitional Evidence* fails, it is possible to construct diachronic Dutch books for conditionalizing plans.[16] In fact, using *Corollary 1* and *Proposition 3*, we can prove:

Corollary 2 *For any inquiry* $\langle W, E \rangle$ *and any rational prior credence function p, let U be a conditionalizing plan defined on that inquiry relative to p. If* $\langle W, E \rangle$ *is non-partitional, then U is subject to a diachronic Dutch book.*

[16]For this observation, Hild (1998b), Williamson (2000, §10.6), Bronfman (2014), and Gallow (2019).

In other words, when *Partitional Evidence* fails, an agent who complies with a conditionalizing plan will be vulnerable to a diachronic Dutch book. If such Dutch books are irrationality-indicating, then *Plan Conditionalization* may not be true after all. In the rest of this section, I will show that, if *Evidence Externalism* is true, then *Partitional Evidence* will indeed fail, and will make conditionalizing plans vulnerable to diachronic Dutch books.

4.1 | Externalism and Partitional Evidence

The evidence externalist is committed to two claims. The first is the claim that an agent's evidence is either a proposition or a set of propositions. The second is the claim that this proposition or set of propositions entails non-trivial propositions not only about the agent's non-factive mental states (e.g., her phenomenal states) but also about the external world. Typically, evidence externalists take factive mental states, e.g., my *seeing that* there's a hand before me, to be sources of conclusive evidence about states of the external world.

My claim is this. If *Evidence Externalism* is true, then we must either reject *Factivity* or *Negative Introspection*. This is because our mechanisms for gathering evidence about the external world are fallible: sometimes, they give us false information without giving us any clue that this has happened. A wall may look red to me, even though it is white and lit up with trick red lighting that will make any surface look red. If my evidence before an inquiry does not rule out the possibility that I might find myself in such a scenario as a result of my inquiry, then both *Factivity* and *Negative Introspection* cannot be true. To see why, let's extend this example. Suppose *Evidence Externalism* is true, and I know all the relevant epistemological facts. Now, consider:

> *Red Wall*. I am about to enter a room where I will face a wall. I have a rational credence of 0.99 that the wall will be red. My evidence entails that, if the wall is red, I will see that the wall is red. But I cannot rule out—and assign a rational credence of 0.01 to—the possibility that the wall will be white but lit up with red light. My evidence entails that, if that happens, the wall will look red to me.

Here, I assign non-zero credence to the possibility that the wall is white but lit up with red light. If *Factivity* is true, then, by lights of my prior evidence, my future evidence only entails truths. So, in that scenario, my future evidence will not entail that the wall is red. But, in that scenario, since the wall looks red to me and I have no reason to suspect that anything is amiss, I will not be able to rule out the possibility that I see that the wall is red. If (by my own lights) seeing that the wall is red suffices for me to have conclusive evidence for the claim that the wall is red, then my future evidence in that scenario will not rule out the possibility that my future evidence entails that the wall is red. So, *Negative Introspection* will fail.

The argument, more generally, is this. Suppose, before an inquiry, the agent anticipates that, in the course of that inquiry, she will gain a piece of evidence that entails a non-trivial proposition P about the external world. We can create a similar inquiry where the agent not only anticipates being in that scenario, but also cannot rule out the possibility of being in a phenomenally indistinguishable scenario in which P is false but she gains the same information from the same source of information without having any clue that P is false. In this latter possibility, the agent will not be able to rule out the possibility that her evidence entails P. So, if *Factivity* is true, then the agent's evidence will not entail P in such a scenario, but also will not entail that it does not entail P. Thus, *Negative Introspection* will be false. With respect to cases of this sort, therefore, we must either reject *Factivity* or *Negative Introspection*.

TABLE 1 Payoffs of B_{old} and B_{new}

	Red	***~Red***
B_{old}	0	1
B_{new}	0	-1

An evidence externalist need not reject *Positive Introspection*.[17] Moreover, even if she does, she may acknowledge that, in cases like *Red Wall*, our inquiries will often satisfy this constraint. Here, I shall be concerned with such cases.

4.2 | Consequences

Suppose *Factivity* fails in *Red Wall*. Then, after I enter the room, if I face a white wall lit up with red light, my evidence will entail that the wall is red. Let's represent this inquiry using the simple structure $\langle W, E \rangle$, where W contains just two worlds r and w: r is the world where the wall is red, and w the world where the wall is white. In any world in W, the strongest evidence I will gain is that the wall is red. Call this proposition *Red*, i.e., the singleton set containing the world r. So, $E(r) = E(w) = Red$. We can depict this structure with Figure 1.

If p is my rational prior credence function before entering the room, my prior credence in *Red* will be $p(Red) = 0.99$. If I update by conditionalizing on *Red*, then my posterior credence in *Red* should be $p(Red|Red) = 1$. Since my posterior credence in *Red* is uniformly higher than my prior credence in both r and w, the conditionalizing plan will not be reflective in this case.

Suppose I commit myself to this plan. Then, by my own lights, I will be vulnerable to an *invariant* diachronic Dutch book. To see this, consider two books B_{old} and B_{new} defined on $\{Red, \sim Red\}$, such that B_{old} is sold to me at maximum price before I enter the room, and B_{new} is sold to me at maximum price after I enter the room irrespective of the wall's colour. The payoffs of these books are given in Table 1.

Since, before I enter the room, my credence in $\sim Red$ is 0.01, the maximum price I can rationally pay for B_{old} is $0.01 \times 1 = 0.01$. Suppose I pay this. But then, after I enter the room, my credence in $\sim Red$ drops to 0. So, the maximum price I'd be rational to pay for B_{new} is $-1 \times 0 = 0$. So, I will rationally accept the bet by paying nothing. But note that, if I accept both books, the net payoff in both r and w will be 0. Thus, in both cases, I will lose 0.01.

Suppose *Factivity* is true in *Red Wall*. This situation can be represented by Figure 2. Here, *Negative Introspection* fails: even though my future evidence in r entails *Red*, my evidence in w does not. In fact, $E(w) = \{r, w\}$. So, if I update by conditionalization, my posterior credence in *Red* in r will rise to 1, but, in w, it will match my prior credence in *Red*. In this case, the conditionalizing plan will be reflective. But it will not be transparent. For, in w, even though the strongest evidence I gain is $\{r, w\}$,

[17]The evidence externalist could reject *Positive Introspection*. If we think that a piece of information can have the status of evidence only if it is safely or reliably acquired from some information-gathering mechanism, then we can run Williamson's (2000, ch. 5) anti-KK argument against *Positive Introspection*. The basic premise will be that even if an agent safely acquires a piece of information, she may not be able to safely determine that it is safely acquired, so she may not have evidence that that piece of information has the status of evidence. However, Williamson's anti-KK argument depends on the assumption that an agent can know certain controversial margin-for-error principles. This assumption has been rejected by others such as Greco (2014), Stalnaker (2015), and Das and Salow (2018).

TABLE 2 Payoffs of B^*_{old}, B_r and B_w

	Red	*~Red*
B^*_{old}	1	1.99
B_r	0	0
B_w	0	-1

I will not be certain that the strongest evidence I have gained is $\{r, w\}$. I will be highly confident that the strongest evidence I have gained is *Red*.

Suppose I commit myself to this conditionalizing plan. Then, by my own lights, I will be vulnerable to a *variant* diachronic Dutch book. Suppose B^*_{old} is a book that is sold to me at maximum price before I enter the room. Suppose B_r is a book that is sold to me at maximum price after I enter the room in r, and B_w is a book that is sold to me at maximum price after I enter the room in w. The payoffs of these books are given in Table 2.

The maximum price I can rationally pay for B^*_{old} is $0.99 \times 1 + 0.01 \times (1.99) = 0.99 + 0.0199 = 1.0099$. In r, since I am certain in *Red*, the maximum price I can rationally pay for B_r is 0. So, in r, the net profit I can make from buying both these books at maximum price is $1 - 1.0099 = -0.0099$. But things are different in w. Since my posterior credence in ~ *Red* is 0.01 in w, the maximum price I can rationally pay for B_w is $(0.01 \times (-1)) = -0.01$. But the net profit I will make from accepting both books will be $1.99 - 1.0099 - 1 - (-0.01) = 1 - 1.0099 = -0.0099$. So, I will lose -0.0099 no matter what happens.

The upshot is this. If we hold *Positive Introspection* fixed in a case like *Red Wall*, then a failure of either *Factivity* or *Negative Introspection* exposes conditionalizing plans to diachronic Dutch books.

4.3 | A Diagnosis

We can generalize this observation.

Suppose we are evidence externalists, and want to reject *Factivity* instead of *Negative Introspection* in cases like *Red Wall*. Then, we can set up an inquiry, where the agent has both positive and negative introspective access to her posterior evidence but her posterior evidence entails falsehoods. In many such cases, an agent's inquiry will satisfy a constraint called *seriality*: an inquiry $\langle W, E \rangle$ is *serial* just in case, for any world w in W, there exists some w^* in $E(w)$. Seriality rules out the possibility that the strongest evidence that an agent gains in an inquiry contradicts the evidence she earlier had. If *Factivity* fails, then an agent can *sometimes*, but not *always*, gain such evidence. So, an agent's inquiry will often satisfy seriality. Here, we can show:

> **Proposition 4** *For any serial, transitive, and euclidean inquiry $\langle W, E \rangle$, let U be a conditionalizing plan defined on that inquiry relative to any rational prior credence function p. Then, the following two claims are equivalent.*
>
> - $\langle W, E \rangle$ is reflexive.
> - U is reflective relative to p.

This shows the following. Suppose, for some inquiry, an agent adopts a conditionalizing plan. Then, if the agent has perfect introspective access to her own posterior evidence in that inquiry and the inquiry satisfies seriality, failures of *Factivity* in that inquiry will be necessary and sufficient for the plan to be non-reflective. *Proposition 2* predicts that complying with such a plan will expose the agent

to an invariant diachronic Dutch book. This explains why I am vulnerable to an invariant diachronic Dutch book in the version of *Red Wall* where *Factivity* fails.

Suppose we are evidence externalists, but we want to preserve *Factivity* by rejecting *Negative Introspection* in cases like *Red Wall*. We can show the following.

> **Proposition 5** *For any inquiry* $\langle W, E \rangle$*, let* U *be a conditionalizing plan defined on* $\langle W, E \rangle$ *relative to a rational prior credence function* p*. Then, the following two claims are equivalent.*
>
> - $\langle W, E \rangle$ is transitive and euclidean.
> - U is transparent.

What does this imply? Suppose, for some inquiry, an agent adopts a conditionalizing plan. Then, failures of either *Positive* or *Negative Introspection* in that inquiry will be necessary and sufficient for the plan to be non-transparent. *Proposition 2* predicts that complying with such a plan will expose the agent to a variant diachronic Dutch book.

These results show that there is a tension amongst *Plan Conditionalization*, *Evidence Externalism*, and *Exploitability*. As we have seen, an evidence externalist should reject either *Factivity* or *Negative Introspection*, and sometimes perhaps even *Positive Introspection*. According to *Proposition 4*, if *Positive* and *Negative Introspection* are held fixed, then failures of *Factivity* (in serial inquiries) make room for invariant diachronic Dutch books for conditionalizing plans. According to *Proposition 5*, failures of *Positive* and *Negative Introspection* expose to conditionalizing plans to variant diachronic Dutch books. So, if *Exploitability* is true, then *Evidence Externalism* and *Plan Conditionalization* cannot be true together.

In what follows, I argue that this tension can only be resolved by giving up the idea that updating plans can be justified (or vindicated) by appealing to diachronic Dutch books.

5 | STRATEGY 1: WEAKENING EXPLOITABILITY

In response to the tension amongst *Plan Conditionalization*, *Evidence Externalism*, and *Exploitability*, one might claim that the fault lies not with *Evidence Externalism* or *Plan Conditionalization*, but rather with *Exploitability*. But, in order to reject *Exploitability*, one does not have to concede that foreseeable exploitability never indicates irrationality. One only has to show that susceptibility to certain kinds of diachronic Dutch books doesn't indicate irrationality.

Let me explain. In the last section, I pointed out that, if *Negative Introspection* fails in *Red Wall*, then the conditionalizing plan will be non-transparent, and therefore will be subject to a variant diachronic Dutch book. But variant diachronic Dutch books that exploit such non-transparency suffer from a flaw: they involve offering the agent different books depending on her evidence. In *Red Wall*, for example, I am offered B_r in the world r where the wall is red, and B_w in the world w where the wall is white.

Why is this a problem? First, an agent's vulnerability to a diachronic Dutch book only indicates the irrationality of her updating plan when she is *foreseeably* exploitable, i.e., if she can see the Dutch book coming but can still be made to lose utility. Suppose I foresee that the bookie will offer me B_r when I am in r and B_w when I am in w. But when she does offer me B_w when I am in w, this will give me information about my epistemic situation. Since I have reason to think that I am in r, I will expect to be offered B_r. So, when I am offered B_w, this will give me evidence that I am in w. Then, if I conditionalize on that evidence, I will just turn down the offer.

Second, if the bookie is to offer me these books depending on my evidence, she must have information about what evidence I have. But in this case, I do not have access to my own evidence. If the bookie does have more information about my evidence than I do, the bookie will have some epistemic

TABLE 3 Payoffs of B^*_{old} and B_w

	Red	*~Red*
B^*_{old}	1	1.99
B_w	0	-1

advantage over me, and therefore will be able to exploit my lack of information about my evidential situation. But, one might argue (as Lewis 1999 does), a diachronic Dutch book that exploits the agent's lack of information in this way cannot indicate her irrationality. This implies that variant diachronic Dutch books that exploit the non-transparency of updating plans cannot indicate irrationality.

For these reasons, we might think that the *variant* diachronic Dutch book that we used to exploit the non-transparency of my conditionalizing plan does not indicate the irrationality of that updating plan. Accordingly, we might want to weaken *Exploitability* while holding on to the idea that some diachronic Dutch books can indicate the irrationality of updating plans. So, consider the following constraint.

> *Invariant Exploitability*. For any inquiry $\langle W, E \rangle$, let U be an updating plan defined on that inquiry. It is rationally permissible for the agent to adopt U before her inquiry only if U is not subject to an invariant diachronic Dutch book.

Modifying *Exploitability* does not help: even if we adopt *Invariant Exploitability*, we can easily show that conditionalizing plans in cases like *Red Wall* may still be subject to an invariant diachronic Dutch book.

Consider a modified version of the variant diachronic Dutch book from §4.2. This consists of two books B^*_{old} and B_w defined on $\{Red, \sim Red\}$ (instead of three), such that B^*_{old} is sold to me at maximum price before I enter the room, and B_w is sold to me at maximum price after I enter the room *irrespective of the colour of the wall*. The payoffs of these books are given in Table 3.

Once again, the maximum price I can rationally pay for B^*_{old} is $0.99 \times 1 + 0.01 \times (1.99) = 0.99 + 0.0199 = 1.0099$. In r, since I am certain in *Red*, the maximum price I can rationally pay for B_w is 0. So, in r, the net profit I can make from buying both these books at maximum price is $1 - 1.0099 = -0.0099$. Since my posterior credence in $\sim Red$ is 0.01 in w, the maximum price I can rationally pay for B_w is $(0.01 \times (-1)) = -0.01$. But the net profit I will make from accepting both books will be $1.99 - 1.0099 - 1 - (-0.01) = 1 - 1.0099 = -0.0099$. So, I will lose 0.0099 no matter what happens. This is an invariant diachronic Dutch book.

This observation generalizes.

Proposition 6 *inquiry $\langle W, E \rangle$, let U be a conditionalizing plan defined on $\langle W, E \rangle$ relative to a rational prior credence function p. Then, the following two claims are equivalent.*

- $\langle W, E \rangle$ is euclidean.
- U is not subject to an invariant diachronic Dutch book.

The upshot: when both *Factivity* and *Positive Introspection* are true but *Negative Introspection* fails, complying with a conditionalizing plan will make an agent vulnerable to an invariant diachronic Dutch book.

According to *Proposition 4*, if *Positive* and *Negative Introspection* are held fixed, then failures of *Negative Introspection* (in serial inquiries) expose to conditionalizing plans to *invariant* diachronic Dutch books. According to *Proposition 6*, if *Factivity* and *Positive Introspection* are held fixed, then failures of *Negative Introspection* make conditionalizing plans vulnerable to *invariant* diachronic

Dutch books. This shows us that, if *Invariant Exploitability* is true, *Evidence Externalism* and *Plan Conditionalization* cannot be true together.

6 | STRATEGY 2: REJECTING EXTERNALISM

Here is a tempting thought. Since *Evidence Externalism* is partly responsible for the tension between *Factivity* and *Negative Introspection*, rejecting it might allow us to justify *Plan Conditionalization* by appealing to diachronic Dutch books. For instance, a Cartesian conception of evidence—on which an agent's evidence only entails propositions about her phenomenal states—could help here. On some versions of this view, our evidence consists of facts we know or are in a position to know by introspection about our phenomenal states. And, arguably, if we do not know something by introspection, we know or are in a position to know that we do not know it. So, these versions of the Cartesian view may preserve both *Factivity* and *Negative Introspection*.

However, such a Cartesian view will face two challenges. First, it is not obvious that the conflict between *Factivity* and *Negative Introspection* can be avoided on this view. For it is not clear that, whenever we fail to know a proposition P about our phenomenal states, we also know (or are in a position to know) that we do not know P. Consider someone who expects to be burnt by hot water. So, when she places her hand under ice-cold water, she may falsely judge for a moment that she is undergoing a hot sensation. At that time, she does not know that she is undergoing a hot sensation. But, since she judges that she is undergoing a hot sensation, she may take herself to know this, and thus may not be in a position to know that she does not know that she is undergoing a hot sensation. According to the Cartesian view we are considering, in this case, the agent's evidence does not entail that she is undergoing a hot sensation, but her evidence also does not entail that her evidence does not entail it. So, *Negative Introspection* fails. In such cases, conditionalizing plans will be subject to invariant diachronic Dutch books.

Second, if a Cartesian view is combined with the idea that we should update by conditionalization, we get sceptical consequences.[18] Imagine an infant who is undergoing her first experiences. If she undergoes a veridical perceptual experience as of there being a hand before her, is it rational for her to be confident that there is a material object of that shape before her? It seems so. But if the Cartesian view is correct, our evidence is exhausted by facts solely about our phenomenal states. If we are required by rationality to update by conditionalization, the infant can only be rationally confident in the proposition M that there is a material object of a certain shape before her if her prior conditional credence in M given her evidence E is much higher than her prior conditional credence in $\sim M$ given her evidence E. This means that her prior credence function must assign a much higher credence to $M \cap E$ than to $\sim M \cap E$. In other words, the agent must assign very low prior credence to sceptical hypotheses on which, even though a material object of a certain hand-like shape appears to her, there is not an object of that shape before her. But since the agent has no prior empirical evidence at this point, she can only assign such low credence to sceptical hypotheses if she has *a priori* reasons for doing so. But it is unclear if we could have *a priori* reasons for discounting contingent sceptical hypotheses.[19] If we cannot have such reasons, the Cartesian view leads to scepticism.

[18]See Neta (2009) for this argument.

[19]White (2006) accepts the view that we can have *a priori* reasons for discounting sceptical possibilities. This commits him to a really strong form of rationalism. Wright (2004) avoids this by claiming that we are entitled to dismiss sceptical possibilities without evidence. This compels him to reject a widely accepted evidentialist conception of epistemic rationality on which we can be rational to believe certain propositions only if we have sufficient evidence for them. Both these views are costly in their own ways.

7 | STRATEGY 3: REJECTING PLAN CONDITIONALIZATION

Given that rejecting *Evidence Externalism* does not help us defend *Plan Conditionalization* using DDBAs, we may reject *Plan Conditionalization*. If conditionalizing plans are not rational for us to adopt, which plans are? The answer, I argue, is not obvious.

7.1 | Failures of Factivity

Start with a version of *Red Wall*, where *Positive* and *Negative Introspection* hold and therefore *Factivity* fails. In the world *r* where the wall is red and in the world *w* where the wall is white but lit up with red light, my posterior evidence entails *Red* (i.e., the proposition that the wall is red). If I update by conditionalizing, my credence in *Red* should rise to 1. This would make the conditionalizing plan non-reflective. So, I will be subject to an invariant diachronic Dutch book (according to *Proposition 4*). What's the way out?

I can avoid this Dutch book only if I do not change my prior credence in *Red* after I have received the new evidence. This is precisely the advice that meta-conditionalizing plans give me. If I comply with a meta-conditionalizing plan here, I will not be conditionalizing on *Red*, but rather on $\boldsymbol{P}$, i.e., the proposition that the strongest evidence I've gained is $\boldsymbol{P}$. This is just ^{Red}. But if I conditionalize on ^{Red}, my prior credence in *Red* will remain fixed at 0.99 even when my posterior evidence entails *Red*.

This way of updating my credences conflicts with:

> *The Entailment-Support Principle*. If an agent's evidence entails $\boldsymbol{P}$, then her evidence conclusively supports $\boldsymbol{P}$.

In both the scenario where the wall is red and the scenario where it is white but lit up with red light, my evidence entails *Red*. So, according to the *Entailment-Support Principle*, my evidence conclusively supports *Red*. Plausibly, epistemic rationality requires me to be certain in whatever my evidence conclusively supports.[20] But this is not the advice that meta-conditionalizing plans give us.

We might try to block this argument by rejecting the *Entailment-Support Principle*. Here are two different kinds of counterexamples. If I see that there's a hawk on the fence post, my evidence (on an externalist picture) may include that information. But suppose I am rationally unsure about whether I really see this: perhaps, hawks are extremely rare in this region, so I cannot believe my eyes! So, I am rationally uncertain about what my evidence is. In such a case, one might think, I cannot be rationally certain that

[20]One might dispute this idea. But here are two counterarguments. First of all, there is an obvious evidentialist rationale for thinking this. Plausibly, in a finite possibility spaces, an agent's rational credences should match the degrees of evidential support that various propositions enjoy. Therefore, under those circumstances, a proposition deserves non-zero credence only if it enjoys some positive degree of evidential support. But if a proposition *P* is conclusively supported by an agent's evidence, then ~ *P* gets no positive degree of evidential support. So, it deserves zero credence. Second, there's a sense in which an agent who violates this principle adopts a set of credences that is both accuracy-wise and practically suboptimal by her own lights. For example, in *Red Wall*, if I assign 0.99 to *Red* and 0.01 credence to ~ *Red* even when my evidence conclusively entails *Red*, my credence function is accuracy-dominated (according to any plausible measure of accuracy) by another credence function that assigns 1 to *Red* and 0 to ~ *Red* relative to the set of worlds that her own evidence does not rule out. Similarly, if I assign 0.01 credence to ~*Red* even when my evidence entails *Red*, then a bookie can sell me at maximum price (of 0.01) a bet that pays 1 unit of utility when ~ *Red* is true and nothing else otherwise, and subject me to sure loss of utility by lights of my evidence. If we think that an agent should adopt credence functions that are accuracy-wise or practically optimal by lights of her evidence, then, in this case, I should indeed be certain in what my evidence conclusively supports.

there is a hawk on the fence post. Similarly, consider a case where I have competently deduced a conclusion from a set of premises, but I have received misleading evidence for thinking that my capacity for deductive reasoning is impaired in some way. Perhaps, I have discovered that I was slipped a reason-distorting drug that is effective in 90 per cent of the cases. But, luckily, unbeknownst to me, it has not worked in this case. Here, I might not be uncertain about what my evidence is, but rather about my capacity for assessing the evidence. One might think that, in such a case, even though my evidence may entail that the conclusion I have drawn (assuming that the premises are entailed by my evidence), I cannot be rationally certain in the conclusion.[21] Such cases strongly suggest that the *Entailment-Support Principle* is false.

I have two responses. First, even if these judgements about these cases are correct, it seems clear that *Red Wall* is not a case of either of these kinds. In the version of *Red Wall* where *Factivity* fails, when my evidence comes to entail ***Red***, my evidence entails that it entails ***Red***, it entails that it entails that it entails ***Red***, and so on. Thus, I have perfect access to what my evidence entails. But I lack that kind of access in the example of the hawk on the fence post. Moreover, I may also have no reason for thinking that my capacity for reasoning (or, generally, for rationally assessing what my evidence entails) is impaired in any way. But I have such evidence in the second case involving competent deduction. What this shows that, even if the unqualified *Entailment-Support Principle* is false, a qualified version of the principle (which sets aside cases of these two kinds) will remain plausible. That will be enough for our purposes.

Here is my second response. If we reject all (qualified or unqualified) versions of the *Entailment-Support Principle* in a case like *Red Wall*, we end up with a form of scepticism. Suppose we reject the *Entailment-Support Principle* by accepting a meta-conditionalizing plan, or, more generally, an updating plan that requires us not to raise our credence in ***Red*** in *Red Wall*. Such a plan will be *sceptic-friendly*: it will recommend that we not raise our credence in a proposition about the external world whenever we assign non-zero credence to a sceptical possibility where, unbeknownst to us, we are misled about that proposition. If we accept *Evidence Externalism* in order to avoid external world scepticism, it will be counterproductive for us to accept such a plan. For, by the reasoning discussed in the last section, if we were to comply with such a plan, we would never justifiably believe anything about the external world.

7.2 | Failures of Negative Introspection

Next, consider the version of *Red Wall* where *Factivity* is true but *Negative Introspection* fails. Given that *Factivity* is true, I should gain no new evidence in w, but my future evidence should entail ***Red*** in r. In r, I have perfect access to what my evidence entails, and have no evidence that suggests that I am incapable of assessing my evidence correctly. Thus, according to a suitably qualified version of the *Entailment-Support Principle*, my evidence should conclusively support ***Red*** in r. So, my posterior credence in ***Red*** in r should be 1. In w, if I retain my prior credence in ***Red***, my updating plan will be subject to an invariant diachronic Dutch book (according to *Proposition 6*). If I increase my credence in ***Red*** in w, my updating plan will be biased in favour of ***Red***, and therefore will not be reflective. So, again, it will be subject to an invariant diachronic Dutch book (according to *Proposition 2*). What's the way out?

I can avoid an invariant Dutch book in this case by decreasing my credence in ***Red*** in w. A meta-conditionalizing plan will give me precisely this advice. If I comply with a meta-conditionalizing plan, I will not be conditionalizing on W in w, but rather on $[\mathbf{E} = W]$, which is just the singleton set

[21]For seminal discussion of cases of defeat involving these two kinds of higher-order evidence, see Lasonen-Aarnio (2010) and Christensen (2010).

~ *Red* containing *w*. But if I conditionalize on ~ *Red*, my prior credence in *Red* will drop to 0 even though I gainno new evidence against *Red* in *w*.

This seems quite strange. This implies: in cases of this sort, the only updating plans that are immune from invariant diachronic Dutch books are ones that require the relevant agent to change her credences even though her evidence does not change as a result of her inquiry. This violates:

> *The Fixity Principle*. If an updating plan is epistemically rational for an agent to adopt before an inquiry, then it does not require her to change her credences as a result of that inquiry unless her posterior evidence is different from her prior evidence.

The *Fixity Principle* seems quite plausible.

However, one could argue that *Fixity Principle* does not hold in cases where an agent's epistemic standards, i.e., her standards for weighing evidence, change without any change in her evidence.[22] To adapt an example from Cohen (2000), I may be rationally certain that, if I go on to study philosophy at Harvard, my epistemic standards for assessing certain pieces of philosophical evidence—e.g., intuitions that support the analytic-synthetic distinction—will change even though all the relevant evidence I have on the matter will not. In a slightly idealized version of such a scenario, I may rationally adopt an updating plan that allows me to assign a different posterior credence to the proposition that there is a genuine analytic-synthetic distinction even though my overall evidence does not change. Even if this is true, this does not affect our argument about *Red Wall*. There is no reason for us to think that, in that scenario, my epistemic standards change after I enter the room and look at the wall. So, rejecting the *Fixity Principle* will not help here. As long as a qualified version of the *Fixity Principle* is true, we will be able to run a version of our argument once more.

A different strategy of rejecting this argument will be to say that I gain some evidence about the colour of the wall when I enter the room in *w*. This seems implausible. Even before entering the room, I may know how the wall will look to me when I enter the room, e.g., by watching a video captured by someone who entered the room just before me. So, when I enter the room and the wall looks exactly that way to me, I presumably shouldn't get any new evidence about the appearance of the wall. Moreover, if the wall isn't in fact red but only lit up with trick red light, then it's unclear how I could get any new evidence about the colour of the wall in this case.[23]

[22]Such a view will follow from a form of *intrapersonal permissivism*, i.e., the view that it is rationally permissible for an agent to assign different credences to the same proposition relative to the same body of evidence (at different times). For discussion of the notion of epistemic standards, see White (2005), Schoenfield (2014), and Kopec and Titelbaum (2019).

[23]Another strategy might be to say that, on entering the room, I gain *self-locating information* about where I am and what time it is, i.e., information that is not about impersonal or objective subject-matters like the colour of the wall or its appearance. For defences of the possibility of such self-locating information, see Lewis (1979) and Perry (1979). For example, on entering the room, I may learn that I am now here in the room. But it is unclear why that kind of self-locating information should boost my credence in a non-self-locating proposition about the colour of the wall, i.e., the proposition that the wall is not red. While there are other cases where such self-locating information seems to rationally impact credences about non-self-locating propositions, this happens (at least one diagnosis) because, in such cases, the agent loses track of who she is or where she is or what time it is. The Sleeping Beauty puzzle is the best example of this phenomenon; for the example, see Elga (2000). Halpern (2006) offers the diagnosis that, in cases like this, rational changes in the agent's doxastic attitudes towards non-self-locating propositions are due to what he calls *asynchrony*, i.e., the agent's having imperfect information about what time it is. Importantly for us, nothing of this sort happens in *Red Wall*. Therefore, someone who adopts this strategy owes us an explanation of why I should adopt an updating plan that recommends changing my credences in *Red* and ~ *Red* in response to apparently irrelevant self-locating information.

There is a more general lesson. If suitably qualified versions of the *Entailment-Support Principle* and the *Fixity Principle* are true, then any updating plan that is rational for us to adopt in cases like *Red Wall* will be subject to an invariant diachronic Dutch book. Therefore, anyone who wishes to justify or vindicate updating plans by appealing to diachronic Dutch books must reject plausible principles about evidential support or epistemic rationality.

8 | CONCLUSION

DDBAs are motivated by principles like *Exploitability*. These are motivated by instrumentalism, i.e., the view that we can justify or vindicate norms of epistemic rationality by appealing to principles of instrumental rationality. The upshot of my argument is this. Given the tension amongst *Invariant Exploitability*, *Evidence Externalism*, and *Plan Conditionalization*, the defenders of DDBAs must reject either *Evidence Externalism* or *Plan Conditionalization*. If they reject *Evidence Externalism*, it is not obvious that they can block external world scepticism or diachronic Dutch books from arising for conditionalizing plans. If they reject *Plan Conditionalization*, then they are forced to violate plausible principles about evidential support or epistemic rationality. So, we cannot rely on DDBAs for the purpose of justifying or vindicating updating plans.

This argument generalizes. Using the results proved by Schoenfield (2017), we can show that there is a similar tension amongst *Evidence Externalism*, *Plan Conditionalization*, and the principle that an agent is rational to adopt only those updating plans which maximize expected accuracy. The same arguments that I have given here would show that this tension cannot be easily resolved by rejecting either *Evidence Externalism* or *Plan Conditionalization*. This should not only cast doubt on accuracy-based arguments for norms governing updating plans, but should also give us reason to be sceptical about instrumentalist defences of updating plans.

Proofs *Proof of Proposition 1*. Let $\langle W, E \rangle$ be any inquiry and p be any rational prior credence function defined on subsets of W. Let U be a probabilistic updating plan defined on that inquiry. We want to show that the following claims are equivalent.

- U is transparent and reflective to p.
- U is a meta-conditionalizing plan relative to p.

This proof is divided into two parts.

First, we want to show that, if U is a meta-conditionalizing plan, then U is both reflective and transparent. Suppose U is a meta-conditionalzing plan. Let the set of the strongest pieces of evidence that an agent could get be $\mathscr{E} = \{E_1, E_2, \ldots, E_k\}$. So, for any E_i, $U_i = p(. \mid [\mathbf{E} = E_i])$ is the credence function that U recommends relative to E_i. So, U is transparent. Moreover, by the law of total probability, for any proposition P, $p(P) = \sum_{E_i \in \varepsilon} p(P \mid [\mathbf{E} = E_i])p([\mathbf{E} = E_i])$. This means that U is reflective.

Next, we want to show that, if U is both reflective and transparent, then U is a meta-conditionalizing plan. Suppose that U is both transparent and reflective. Since U is reflective, there exists a probability function p^*, such that, for any $E_m \in \mathscr{E}$,

$$p([\mathbf{E} = E_m]) = \sum_{i=1}^{k} U_i([\mathbf{E} = E_m])p^*([\mathbf{E} = E_i]).$$

TABLE 4 Payoffs of $B, B_{[\mathbf{E}=E_i]}$ and $B_{\sim[\mathbf{E}=E_i]}$.

	$[\mathbf{E}=E_i]$	$\sim[\mathbf{E}=E_i]$
B	$1+U_i(\sim[\mathbf{E}=E_i])$	1
$B_{[\mathbf{E}=E_i]}$	-1	0
$B_{\sim[\mathbf{E}=E_i]}$	0	0

Since U is transparent, $U_i(\ [\mathbf{E}=E_m]) = 1$ if $i=m$; otherwise, $U_i([\mathbf{E}=E_m]) = 0$. Therefore,

$$p([\mathbf{E}=E_m]) = p^*([\mathbf{E}=E_m]).$$

This implies that, for any proposition P,

$$p(P \cap [\mathbf{E}=E_m]) = \sum_{i=1}^{k} U_i(P \cap [\mathbf{E}=E_m])p([\mathbf{E}=E_i]).$$

Since U is transparent and probabilistic, for any i, if $i \neq m$, $U_i([\mathbf{E}=E_m]) = 0$. Moreover, for any i, if $i=m$, $U_i([\mathbf{E}=E_m]) = U_m(P)$. So, we have:

$$p(P \cap [\mathbf{E}=E_m]) = U_m(P)p([\mathbf{E}=E_m]).$$

This just means that U is a meta-conditionalizing plan. QED.

Proof of Proposition 2. Let $\langle W, E\rangle$ be any inquiry and p be any rational prior credence function defined on subsets of W. Let U be a probabilistic updating plan defined on that inquiry. We want to show:

- If U is not transparent, then U is subject to a variant diachronic Dutch book.
- If U is not reflective relative to p, then U is subject to an invariant diachronic Dutch book.
- If U is transparent and reflective relative to p, then U is not subject to a variant diachronic Dutch book.

This proof is divided into three parts.

Part 1. First, we shall show that, if U is not transparent, then it is subject to a variant diachronic Dutch book. If U is not transparent, then there is some evidence-proposition E_i (i.e., one of the outputs of E) such that the credence function recommended by U relative to E_i—call it U_i—assigns a credence of less than 1 to $[\mathbf{E}=E_i]$, i.e., $U_i([\mathbf{E}=E_i]) < 1$. Since U_i is a probability function, $U_i(\sim[\mathbf{E}=E_i]) > 0$.

We can construct three books: $B, B_{[\mathbf{E}=E_i]}$, and $B_{\sim[\mathbf{E}=E_i]}$. The payoffs of these books are given in Table 4.

First, B is sold to the relevant agent before she gathers evidence at the maximum price of:

$$\begin{aligned}\text{price } B, p &= p([\mathbf{E}=E_i])\ 1 + U_i \sim [\mathbf{E}=E_i] + p(\sim[\mathbf{E}=E_i]) \\ &= 1 + p([\mathbf{E}=E_i])U_i \sim [\mathbf{E}=E_i].\end{aligned}$$

Second, $B_{[\mathbf{E}=E_i]}$ is sold to the agent after she has gained E_i as her evidence at the maximum price of $\texttt{price}\left(B_{[\mathbf{E}=E_i]}, U_i\right) = U_i([\mathbf{E}=E_i]) \times (-1) = -U_i([\mathbf{E}=E_i])$. Finally, when the agent gains any

evidence-proposition E_j as her future evidence such that $i \neq j$, $B_{\sim[\mathbf{E}=E_i]}$ is sold to her at the maximum price of $\texttt{price}\left(B_{\sim[\mathbf{E}=E_i]}, U_j\right) = 0$.

In any $[\mathbf{E}=E_i]$-world w, the net profit is:

$$\begin{aligned}
&\texttt{profit}(B,p,w) + \texttt{profit}\left(B_{[\mathbf{E}=E_i]}, U_i, w\right)\\
&= \texttt{payoff}(B,w) - \texttt{price}(B,p)\\
&+ \texttt{payoff}\left(B_{[\mathbf{E}=E_i]}, w\right) - \texttt{price}\left(B_{[\mathbf{E}=E_i]}, U_i\right)\\
&= (1 + U_i(\sim[\mathbf{E}=E_i]) - 1 - p[\mathbf{E}=E_i])\, U_i(\sim[\mathbf{E}=E_i]) + (-1 + U_i(\mathbf{E}=E_i))\\
&= (-p[\mathbf{E}=E_i])\, U_i(\sim[\mathbf{E}=E_i]) < 0.
\end{aligned}$$

In any $\sim[\mathbf{E}=E_i]$-world w, where the evidence that the agent gains is E_j, the net profit is:

$$\begin{aligned}
&\texttt{profit}(B,p,w) + \texttt{profit}\left(B_{\sim[\mathbf{E}=E_i]}, U_j, w\right)\\
&= \texttt{payoff}(B,w) - \texttt{price}(B,p)\\
&+ \texttt{payoff}\left(B_{[\mathbf{E}=E_i]}, w\right) - \texttt{price}\left(B_{\sim[\mathbf{E}=E_i]}, U_j\right)\\
&= (1 - 1 - p(\mathbf{E}=E_i])\, U_i(\sim[\mathbf{E}=E_i]) + (0-0)\\
&= (-p[\mathbf{E}=E_i])\, U_i(\sim[\mathbf{E}=E_i]) < 0.
\end{aligned}$$

So, in either case, the agent loses some amount of utility.

Part 2. Next, we show that, if U is not reflective, then it is subject to an invariant diachronic Dutch book.[24] Since there are $n$worlds in W(for some finite n), we can represent each world w_j in Was an n-dimensional vector, with 1 at the j-th place and 0s elsewhere. Similarly, the prior credence function p as well as each U_i can be represented as a n-dimensional vector, such that the j-th place is occupied by the value that p or the relevant U_i assigns to w_j. Since U is probabilistic but not reflective, then p does not lie in the convex hull of the U_i's. By a well-known Hyperplane Separation Theorem, we can now show that there is an n-dimensional vector B such that $B \cdot (U_i - p) < 0$ for any U_i.[25] Let the vector B be a book defined on the finest partition on W, where the value that occupies the j-th place of the vector is the stake of B on $\{w_j\}$. Then, we can construct another book B^* such that for any $w_j \in W$, $B^*(w_j) = -B(w_j)$. So, for any $w \in W$ where $E(w) = E_i$,

$$\begin{aligned}
&\texttt{profit}(B,p,w) + \texttt{profit}(B^*, U_i, w)\\
&= (B(w) - \sum_{w_j \in W} p(w_j)B(w_j)) + (B*(w) - \sum_{w_j \in W} U_i(w_j)B*(w_j))\\
&= (B(w) - \sum_{w_j \in W} p(w_j)B(w_j)) + (-B(w) - \sum_{w_j \in W} U_i(w_j)(-B(w_j)))\\
&= \sum_{w_j \in W} B(w_j)(U_i(w_j) - p(w_j))\\
&= B \cdot (U_i - p)\\
&< 0.
\end{aligned}$$

[24] A similar proof strategy is used by Paris (2001), Williams (2012), and especially Pettigrew (2020).

[25] Here's a proof of the relevant theorem. Let the convex hull of the U_i's be $X = \{U_1, U_2, \ldots, U_k\}^+$. Since X is closed convex set, there is a unique point a such that a is a point in X that is uniquely closest to p and $a \neq p$. Let $b = \frac{1}{2}(p + a)$ and let H be the plane of vectors y such that $(y - b) \cdot (p - a) = 0$. But then X cannot intersect H: if it did, say, at c, then there would be a point d which would be on the line between c and a in Y and which would be closer to p than a. So, X must lie entirely on the opposite side of H from p. This would mean that, for any U_i, $p \cdot (b - p) < 0 < U_i \cdot (b - p)$. This, in turn, would imply that $(U_i - p) \cdot (p - b) < 0$.

Therefore, U is subject to an invariant diachronic Dutch book.

Part 3. Finally, we want to show that, if U is transparent and reflective, then it is not subject to a diachronic Dutch book. Suppose, for *reductio*, U is subject to a diachronic Dutch book. If it is both transparent and reflective, then (by *Proposition 1*) it is a meta-conditionalizing plan. Since U is subject to a (variant or invariant) diachronic Dutch book, we know that there exist books $B, B_1, \ldots, B_k$ defined on a set of propositions, such that, for any $w \in W$, if $E(w) = E_i$, $\texttt{profit}(B, p, w) + \texttt{profit}\left(B_i, U(w), w\right) < 0$. We show that, if this is so, then p is subject to a synchronic Dutch book B^*, i.e., buying B^* at maximum price relative to p results in sure loss of utility.

For any $w \in W$, if E_i is the evidence-proposition that the agent gets in w, let $B_i = B_w$. So, we define the payoff of B^* as follows.

$$\texttt{payoff}\left(B^*, w\right) = \texttt{payoff}\left(B, w\right) + \texttt{payoff}\left(B_w, w\right) - \texttt{price}\left(B_w, U(w)\right).$$

We can show the maximum price of B^* relative to p is:

$$\begin{aligned}
&\text{price}\left(\mathrm{B}^*, p\right)\\
&= \sum_{w \in W} p(w)\,\text{payoff}(B^*(w))\\
&= \sum_{w \in W} p(w)\,[\,\text{payoff}(B, w) + \text{payoff}(B_w, w) - \text{price}(Bw, U(w))\,]\\
&= \sum_{w \in W} p(w)\,\text{payoff}(B, w) + \sum_{w \in W} p(w)\,\text{payoff}(B_w, w)\\
&\quad - \sum_{w \in W} p(w)\,\text{price}(B_w, U(w))\\
&= \text{price}(B, p) + \sum_{w* \in W} p(w)\,\text{payoff}(B_w, w)\\
&\quad - \sum_{w \in W} p(w) \sum_{w* \in W} p(w^* \mid [\,\mathbf{E} = E(w)\,])\,\text{payoff}(B_w, w^*)\\
&= \text{price}(B, p) + \sum_{w \in W} p(w))\,\text{payoff}(B_w, w)\\
&\quad - \sum_{w \in W} p(w) \sum_{w* \in [\,\mathbf{E} = E(w)\,]} p(w^* \mid [\,\mathbf{E} = E(w)\,])\,\text{payoff}(B_w, w^*)\\
&= \text{price}(B, p) + \sum_{w \in W} p(w)\,\text{payoff}(B_w, w) - \sum_{w \in W} p(w)\,\text{payoff}(B_w, w)\\
&= \text{price}(B, p)
\end{aligned}$$

This means that, for any w,

$$\begin{aligned}
\texttt{profit}\left(B^*, p, w\right), &= \texttt{payoff}\left(B^*, w\right) - \texttt{price}\left(B^*, p\right)\\
&= \texttt{payoff}\left(B, w\right) + \texttt{payoff}\left(B_w, w\right) - \texttt{price}\left(B_w, U(w)\right) - \texttt{price}\left(B, p\right)\\
&= \texttt{profit}\left(B, p, w\right) + \texttt{profit}\left(B_w, U(w), w\right) < 0.
\end{aligned}$$

But this cannot be the case. For, according to the converse Dutch book theorem for probabilism, a probability function cannot be subject to a synchronic Dutch book. QED.

Proof of Proposition 3. Relative to an inquiry $\langle W, E\rangle$, *let* $\mathscr{E} = \{E(w) : w \in W\}$ *be the set of all the strongest pieces of evidence that an agent could gain. We need to prove two conditionals:*

Claim 1 $\langle W, E\rangle$ is reflexive iff, for any proposition $P \in \mathscr{E}$, $[\mathbf{E} = P] \subseteq P$.

Claim 2 $\langle W, E\rangle$ is transitive and euclidean iff, for any proposition $P \in \mathscr{E}$ such that $P \subseteq [\mathbf{E} = P]$.

First, we prove *Claim 1*. Suppose $\langle W, E \rangle$ is reflexive. Then, for any $w \in W$, $w \in E(w)$. Therefore, $[\mathbf{E} = E(w)] \subseteq E(w)$. Now, suppose $\langle W, E \rangle$ is not reflexive. Then, for some $w \in W$, $w \notin E(w)$. But then, it is not the case that $[\mathbf{E} = E(w)] \subseteq E(w)$.

Now, we prove *Claim 2*. Suppose $\langle W, E \rangle$ is transitive and euclidean. For *reductio*, suppose, for some $w \in W$, it is not the case that $E(w) \subseteq [\mathbf{E} = E(w)]$. So, there is some $w^* \in E(w)$, $E(w^*) \neq E(w)$. This can happen in two ways: either there exists some w^{**} in $E(w)$ such that $w^{**} \notin E(w^*)$, or there exists some w^{**} in $E(w^*)$ such that $w^{**} \notin E(w)$. The first possibility is ruled out by euclideanness, while the second possibility is ruled out by transitivity. Therefore, for any $w \in W$, $E(w) \subseteq [\mathbf{E} = E(w)]$. Now, suppose $\langle W, E \rangle$ is either not transitive or not euclidean. If it is not transitive, then there exist two worlds w, w^*, such that $w^* \in E(w)$ but it is not the case that $E(w^*) \subseteq E(w)$. So, it is not the case that $E(w) \subseteq [\mathbf{E} = E(w)]$. If it is not euclidean, there exists a world $w \in W$ such that, for some worlds $w^*, w^{**} \in E(w)$, $w^* \notin E(w^{**})$; in that case, $E(w^{**}) \neq E(w)$. So, it is not the case that $E(w) \subseteq [\mathbf{E} = E(w)]$. QED.

Proof of Proposition 4. For any serial, transitive, and euclidean inquiry $\langle W, E \rangle$, let U be a conditionalizing plan defined on that inquiry relative to any rational prior p. Then, the following two claims are equivalent.

- $\langle W, E \rangle$ is reflexive.
- U is reflective relative to p.

Suppose an inquiry $\langle W, E \rangle$ is serial, transitive and euclidean.

First, we show that, if $\langle W, E \rangle$ is reflexive, then there exists no prior credence function p such that any conditionalizing plan U based on p is non-reflective. Suppose $E_1, E_2, \ldots, E_k$ are the strongest pieces of posterior evidence that the agent could get as a result of her inquiry. If $\langle W, E \rangle$ is reflexive, transitive and euclidean, then, for any i between 1 and k (inclusive), $E_i = \left[\mathbf{E} = E_i\right]$ (by *Proposition 3*). In that case, any conditionalizing plan U based on p is also a meta-conditionalizing plan. From *Proposition 1*, we know that meta-conditionalizing plans are reflective.

Second, we show that, if $\langle W, E \rangle$ is *not* reflexive, then, for any regular probability function p, any conditionalizing plan U based on p is non-reflective. Since $\langle W, E \rangle$ is not reflexive, there exists a world $w \in W$ such that $w \notin E(w)$. Either there is a world $w^* \in W$ such that $w \in E(w^*)$, or there is not. Suppose there is such a world w^*. Now, by seriality, let w^{**} be a world such that $w^{**} \in E(w)$. So, by transitivity, $w^{**} \in E(w^*)$. But then, since $w \in E(w^*)$ and $w^{**} \in E(w^*)$, $w \in E(w^{**})$ by euclideanness. Again, since $w^{**} \in E(w)$ and $w \in E(w^{**})$, by transitivity $w \in E(w)$. But this contradicts our earlier assumption. Therefore, there is no world $w^* \in W$ such that $w \in E(w^*)$. If this is correct, then, for any $E(w^*)$, $p(\sim\{w\} \mid E(w^*)) = 1$. Since p is regular, $p(\sim\{w\}) < 1$. Since U is a conditionalizing plan based p, then, for any w^*, $U(w^*) = p(. \mid E(w^*))$. In that case, U is not reflective. QED.

Proof of Proposition 5. We want to show that, for any inquiry $\langle W, E \rangle$, if U is a conditionalizing plan based on a rational regular prior credence function p, the following two claims are equivalent.

- $\langle W, E \rangle$ is transitive and euclidean.
- U is transparent.

By *Claim 2* in the proof of *Proposition 3*, we know that an inquiry is both euclidean and transitive iff, for any $w \in W$, $E(w) \subseteq [\mathbf{E} = E(w)]$. Suppose $\langle W, E \rangle$ is transitive and euclidean. Then, for any $w \in W$, if $c = U(w)$, $c([\mathbf{E} = E(w)]) = p([\mathbf{E} = E(w)] \mid E(w)) = 1$. So, U is transparent. Now, suppose $\langle W, E \rangle$ is not euclidean and transitive. So, for some $w \in W$, it is not the case that $E(w) \subseteq [\mathbf{E} = E(w)]$. Since p is regular, $p([\mathbf{E} = E(w)] \mid E(w)) < 1$. Thus, U will not be transparent. QED.

TABLE 5 Payoffs of B^{**}_{old} and B^{**}_{new}

	$[\mathbf{E} = E(z)]$	$\sim [\mathbf{E} = E(z)]$
B^{**}_{old}	$1 + p(\sim [\mathbf{E} = E(z)] \mid E(z))$	1
B^{*}_{new}	-1	0

Proof of Proposition 6. We want to show that, for any reflexive and transitive inquiry $\langle W, E \rangle$, if U is a conditionalizing plan defined on $\langle W, E \rangle$ based on a rational prior credence function p, the following two claims are equivalent.

- $\langle W, E \rangle$ is euclidean.
- U is not subject to an invariant diachronic Dutch book.

If $\langle W, E \rangle$ is reflexive, transitive and euclidean, then, by *Proposition 3*, for any $w \in W$, $E(w) = [\mathbf{E} = E(w)]$ In that case, any conditionalizing plan U based on a regular prior credence function p will also be a meta-conditionalizing plan. From *Proposition 2*, we know that such plans aren't subject to diachronic Dutch books. So, all we need to prove here is that, if $\langle W, E \rangle$is reflexive and transitive but not euclidean, U is subject to an invariant diachronic Dutch book.

If $\langle W, E \rangle$ is reflexive and transitive, but not euclidean, then there exist two worlds $x, y \in W$, such that $x \in E(y)$ but $y \notin E(x)$.[26] Now, given the finiteness of W, there is also a world $z \in W$, such that $y \in E(z)$, but there is no world w such that $E(z) \subset E(w)$.

Note two things. First of all, $x \in E(z)$, but $z \notin E(x)$. For, if that were so, then, by transitivity, y would be in $E(x)$. So, $p(\sim [\mathbf{E} = E(z)] \mid E(z)) > 0$. Second, for any world w, if $E(w) \neq E(z)$, there is no world $w^* \notin E(w)$ such that $E(w^*) = E(z)$. For, if that were the case, then, by transitivity, $E(z)$ would be a subset of $E(w)$. But, since it is not the case that $E(z) \subset E(w)$, would have to be $E(z)$. So, we get the result that, for any $w \in W$, if $E(w) \neq E(z)$, then $E(w) \subseteq \sim [\mathbf{E} = E(z)]$. In that case, $p(\sim [\mathbf{E} = E(z)] \mid E(w)) = 1$.

We construct two books B^{**}_{old} and B^{**}_{new} such that B^{**}_{old} is sold to the relevant agent before the inquiry at maximum price, and B^{**}_{new} sold to her after the inquiry at maximum price. The payoffs of these books are given in Table 5.

First, the maximum price that the agent can rationally pay for B^{**}_{old} relative to p is:

$$\begin{aligned} \texttt{price}\left(B^{**}_{old}, p\right) &= p([\mathbf{E} = E(z)])(1 + p(\sim [\mathbf{E} = E(z)] \mid E(z))) + p(: [\mathbf{E} = E(z)]) \\ &= 1 + p([\mathbf{E} = E(z)])\, p(\sim [\mathbf{E} = E(z)] \mid E(z)) \end{aligned}.$$

Second, in any $[\mathbf{E} = E(z)]$/>-world w, the agent will buy B^{**}_{new} at the maximum price of:

$$\begin{aligned} \texttt{price}(B^{**}_{new}, p(. \mid E(z))) &= -1 \times p([\mathbf{E} = E(z)] \mid E(z)) + 0 \times p(\sim [\mathbf{E} = E(z)] \mid E(z)) \\ &= -p([\mathbf{E} = E(z)] \mid E(z)) \end{aligned}.$$

Finally, we know that, for any $\sim [\mathbf{E} = E(z)]$-world w, $p(\sim [\mathbf{E} = E(z)] \mid E(w)) = 1$. So, in any world $\sim [\mathbf{E} = E(z)]$-world w, the maximum price that the agent can pay for B^{**}_{new} is:

$$\texttt{price}(B^{**}_{new}, p(. \mid E(w))) = -1 \times 0 + 0 \times 1 = 0.$$

[26] Why is this true? Since $\langle W, E \rangle$ is not euclidean, there exist three worlds x, y, w, such that $w \in E(y)$ and $x \in E(y)$, but $w \notin E(x)$. But, in that case, it cannot be the case that $y \in E(x)$. For, if that were the case, then, by transitivity, w would be in $E(x)$.

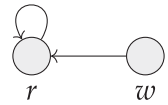

FIGURE 1 A Failure of *Factivity* in *Red Wall*

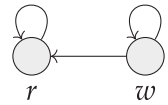

FIGURE 2 A Failure of *Negative Introspection* in *Red Wall*

Therefore, in any $[\mathbf{E} = E(z)]$-world w, the agent's net profit will be:

$$\begin{aligned}
&\texttt{profit}\left(B^{**}_{old}, p, w\right) + \texttt{profit}(B^{**}_{new}, p(\,.\,|E(z)\,), w)\\
&\quad = \left(\texttt{payoff}\left(B^{**}_{old}, w\right) - \texttt{price}\left(B^{**}_{old}, p\right)\right)\\
&\quad + (\texttt{payoff}\left(B^{**}_{new}, w\right) - \texttt{price}(B^{**}_{new}, p(\,.\,|E(z)\,))\\
&\quad = (1 + p(\sim[\mathbf{E} = E(z)]\,|E(z)) - 1 - p([\mathbf{E} = E(z)])\,p(\sim[\mathbf{E} = E(z)]\,|E(z)))\\
&\quad + (-1 + p(\,[\mathbf{E} = E(z)]\,|E(z)\,))\\
&\quad = -p([\mathbf{E} = E(z)])\,p(\sim[\mathbf{E} = E(z)]\,|E(z)\,) < 0.
\end{aligned}$$

And, in any $\sim[\mathbf{E} = E(z)]$-world w, the agent's net profit will be:

$$\begin{aligned}
&\texttt{profit}\left(B^{**}_{old}, p, w\right) + \texttt{profit}(B^{**}_{new}, p(\,.\,|E(w)\,)), w)\\
&\quad = \left(\texttt{payoff}\left(B^{**}_{old}, w\right) - \texttt{price}\left(B^{**}_{old}, p\right)\right)\\
&\quad + (\texttt{payoff}\left(B^{**}_{new}, w\right) - \texttt{price}(B^{**}_{new}, p(\,.\,|E(w)\,))\\
&\quad = (1 - 1 - p([\mathbf{E} = E(z)])\,p(\sim[\mathbf{E} = E(z)]\,|E(z)\,)) + (0 - 0)\\
&\quad = -p([\mathbf{E} = E(z)])\,p(\sim[\mathbf{E} = E(z)]\,|E(z)\,) < 0.
\end{aligned}$$

Thus, the net profit is negative in both cases. QED.[27]

REFERENCES

Briggs, R. A. (2009). Distorted reflection. *Philosophical Review*, *118*(1), 59–85. https://doi.org/10.1215/00318108-2008-029.

Briggs, R. A., & Pettigrew, R. (2020). An accuracy-dominance argument for conditionalization. *Noûs*, *54*(2), 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.

Buchak, L. (2013). *Risk and Rationality*. Oxford: Oxford University Press.

[27]For comments on earlier drafts of this paper, I am grateful to Arif Ahmed, Tim Button, Julien Dutant, Peter Epstein, Rachel Fraser, Richard Holton, Alex Kaiserman, Matthew Mandelkern, Bernhard Salow, Weng Hong Tang, audiences at the Jowett Society at Oxford, the Moral Sciences Club at Cambridge, and the Formal Methods Workshop at KCL, and an anonymous referee and an editor of this journal.

Christensen, D. (1991). Clever bookies and coherent beliefs. *Philosophical Review*, *100*(2), 229–247. https://doi.org/10.2307/2185301.

Christensen, D. (1996). Dutch-book arguments depragmatized: Epistemic consistency for partial believers. *Journal of Philosophy*, *93*(9), 450–479. https://doi.org/10.2307/2940893.

Christensen, D. (2010). Higher-order evidence. *Philosophy and Phenomenological Research*, *81*(1), 185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x.

Cohen, G. A. (2000). *If You're An Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press.

Das, N. (2019). Accuracy and ur-prior conditionalization. *Review of Symbolic Logic*, *12*(1), 62–96. https://doi.org/10.1017/s1755020318000035.

Das, N., & Salow, B. (2018). Transparency and the KK Principle. *Noûs*, *52*(1), 3–23. https://doi.org/10.1111/nous.12158.

deFinetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *17*, 1–68.

Easwaran, K. (2013). Expected accuracy supports conditionalization—and conglomerability and reflection. *Philosophy of Science*, *80*(1), 119–142. https://doi.org/10.1086/668879.

Easwaran, K. (2014). Regularity and hyperreal credences. *Philosophical Review*, *123*(1), 1–41. https://doi.org/10.1215/00318108-2366479.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, *60*(2), 143–147. https://doi.org/10.1111/1467-8284.00215.

Fraassen, B. C. V. (1984). Belief and the will. *Journal of Philosophy*, *81*(5), 235–256. https://doi.org/10.2307/2026388.

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

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. (forthcoming). Updating for externalists. *Noûs*. https://doi.org/10.1111/nous.12307.

Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, *115*(459), 607–632. https://doi.org/10.1093/mind/fzl607.

Greco, D. (2014). Could KK be OK? *Journal of Philosophy*, *111*(4), 169–197. https://doi.org/10.5840/jphil2014111411.

Halpern, J. (2006). Sleeping beauty reconsidered: Conditioning and reflection in asynchronous systems. In Hawthorne, J. and Gendler, T.S., editors, *Oxford Studies in Epistemology, volume1*. Oxford: Oxford University Press.

Hedden, B. (2013). Incoherence without exploitability. *Noûs*, *47*(3), 482–495. https://doi.org/10.1111/j.1468-0068.2011.00842.x.

Hild, M. (1998a). Auto-epistemology and updating. *Philosophical Studies*, *92*(3), 321–361. https://doi.org/10.1023/a:1004229808144.

Hild, M. (1998b). The coherence argument against conditionalization. *Synthese*, *115*(2), 229–258. https://doi.org/10.1023/A:1005082908147.

Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, *65*(4), 575–603. https://doi.org/10.1086/392661.

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

Kemeny, J. G. (1955). Fair bets and inductive probabilities. *Journal of Symbolic Logic*, *20*(3), 263–273. https://doi.org/10.2307/2268222.

Kopec, M., & Titelbaum, M. (2019). When rational reasoners reason differently. In Magdalena Balcerak Jackson & Brendan Balcerak Jackson (Eds.). *Reasoning: New Essays on Theoretical and Practical Thinking*. Oxford: Oxford University Press.

Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives*, *24*(1), 1–21. https://doi.org/10.1111/j.1520-8583.2010.00183.x.

Lehman, R. S. (1955). On confirmation and rational betting. *Journal of Symbolic Logic*, *20*(3), 251–262. https://doi.org/10.2307/2268221.

Leitgeb, H., & Pettigrew, R. (2010a). An objective justification of bayesianism I: Measuring inaccuracy. *Philosophy of Science*, *77*(2), 201–235. https://doi.org/10.1086/651317.

Leitgeb, H., & Pettigrew, R. (2010b). 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. (1979). Attitudes De Dicto and De Se. *Philosophical Review*, *88*(4), 513–543. https://doi.org/10.2307/2184843.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, Vol. *II* (pp. 263–293). Berkeley: University of California Press.

Lewis, D. (1999). Why conditionalize?. *Papers in Metaphysics and Epistemology*, Vol. Cambridge: Cambridge University Press.

Mahtani, A. (2012). Diachronic Dutch book arguments. *Philosophical Review*, *121*(3), 443–450.

Mahtani, A. (2015). Dutch books, coherence, and logical consistency. *Noûs*, *49*(3), 522–537.

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

Neta, R. (2009). Defeating the dogma of defeasibility. In P. Greenough, D. Pritchard, & T. Williamson (Eds.), *Williamson on Knowledge* (pp. 161–82). Oxford: Oxford University Press.

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

Paris, J. B. (2001). A note on the dutch book method. Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications, pages 301–306

Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*(1), 3–21. https://doi.org/10.2307/2214792.

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

Pettigrew, R. (2020). *The Dutch Book Arguments*. Cambridge: Cambridge University Press.

Pettigrew, R. (forthcoming). On the expected utility objection to the Dutch book argument for probabilism. *Noûs*. https://doi.org/10.1111/nous.12286.

Ramsey, F. (1926). Truth and probability. In A. Eagle (Ed.), *Philosophy of Probability: Contemporary Readings* (pp. 52–94). London: Routledge.

Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, *48*(2), 193–218. https://doi.org/10.1111/nous.12006.

Schoenfield, M. (2017). Conditionalization does not maximize expected accuracy. *Mind*, *126*(504), 1155–1187. https://doi.org/10.1093/mind/fzw027.

Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press.

Skyrms, B. (1987). Dynamic coherence and probability kinematics. *Philosophy of Science*, *54*(1), 1–20. https://doi.org/10.1086/289350.

Stalnaker, R. (2015). Luminosity and the KK thesis. In S. C. Goldberg (Ed.), *Externalism, Self-Knowledge, and Skepticism, volume1* (pp. 167–196). Cambridge: Cambridge University Press.

Teller, P. (1973). Conditionalization and observation. *Synthese*, *26*(2), 218–258. https://doi.org/10.1007/bf00873264.

van Fraassen, B. C. (1999). Conditionalization, a new argument for. *Topoi*, *18*(2), 93–96. https://doi.org/10.1023/A:1006286003463.

White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives*, *19*(1), 445–459. https://doi.org/10.1111/j.1520-8583.2005.00069.x.

White, R. (2006). Problems for dogmatism. *Philosophical Studies*, *131*(3), 525–557. https://doi.org/10.1007/s11098-004-7487-9.

Williams, J. R. G. (2012). Generalized probabilism: Dutch books and accuracy domination. *Journal of Philosophical Logic*, *41*(5), 811–840. https://doi.org/10.1007/s10992-011-9192-4.

Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, *31*(2), 131–144. https://doi.org/10.1093/bjps/31.2.131.

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

Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, *67*(3), 173–180. https://doi.org/10.1093/analys/67.3.173.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary*, *78*(1), 167–212. https://doi.org/10.1111/j.0309-7013.2004.00121.x.

Wrónski, L. and Godziszewski, M.T., (2017). The stubborn non-probabilist—'negation incoherence' and a new way to block the Dutch book argument. In A. Baltag J. Seligman & T. Yamada (Eds.). *Logic, Rationality, and Interaction (LORI) 2017*, (pp. 256–267). Berlin and Heidelberg: Springer. https://doi.org/10.1007/978-3-662-55665-8_18

How to cite this article: Das N. Externalism and exploitability. *Philos Phenomenol Res*. 2020;00:1–27. https://doi.org/10.1111/phpr.12742