The Value of Biased Information

Nilanjan Das

Abstract

In this essay, I cast doubt on an apparent truism: namely, that if evidence is available for gathering and use at a negligible cost, then it's always instrumentally rational for us to gather that evidence and use it for making decisions. Call this thesis **Value of Information**. I show that **Value of Information** conflicts with two other plausible theses. The first is the view that an agent's evidence can entail non-trivial propositions about the external world. The second is the view that epistemic rationality requires us to update our credences by conditionalization. These two theses, given some plausible assumptions, make room for rationally biased inquiries where **Value of Information** fails. I go on to argue that this is bad news for defenders of **Value of Information**.

Here is a plausible line of reasoning. Our evidence is our best guide to the truth. To be successful in our theoretical and practical projects, we need to believe the truth about the relevant subject-matters. Therefore, we ought to gather more evidence and use it for making decisions about our projects, unless gathering evidence and using it is too costly. This supports:

> **Value of Information**. Necessarily, when evidence is available to an agent for gathering and use at a negligible cost, it is instrumentally rational for her to gather that evidence and use it for making decisions.[1]

Value of Information might look like a truism. But it isn't.

In this essay, I argue that, given plausible assumptions about instrumental rationality and our sources of information, **Value of Information** is incompatible with two attractive theses.

> **Evidence Externalism**. Any agent's evidence is a proposition or a set of propositions, which can entail non-trivial propositions about the external world.[2]

---

[1]For arguments for this claim, see Peirce [1967], Ramsey [1990], and Good [1967]. Some have shown that Good's argument makes a number of non-trivial assumptions about the preciseness of the agent's credal states, the structure of her future evidence, and the norms of instrumental rationality: see Good [1974], Kadane, Schervish and Seidenfeld [2008], Skyrms [1990], Buchak [2010], Huttegger [2014], Ahmed and Salow [forthcoming] and Dorst [forthcoming].

[2]Prominent defenders of **Evidence Externalism** include McDowell ([1982], [1995]) (at

Conditionalization. If an agent's prior credence function at a time t_1 is p_1 and the strongest evidence she receives between t_1 and a later time t_2 is E, then for any proposition H, her credence at t_2 in H should be:

$$p_2(H) = p_1(H|E) = \frac{p_1(H \cap E)}{p_1(E)} \text{ (provided } p_1(E) > 0).^3$$

My argument involves rationally biased inquiries. Suppose **Evidence Externalism** and **Conditionalization** are true. Then, given plausible assumptions about our sources of information about the external world, an agent can set up an inquiry that is rationally biased in favour of a proposition (in other words, an inquiry that is guaranteed by her own lights to rationally raise her credence in that proposition) (§1). Given a further plausible assumption about instrumental rationality, in such cases, it can be instrumentally irrational for certain agents to gather and use cost-free evidence (§2). I'll argue that this is bad news for defenders of **Value of Information** (§3).

1 Rationally Biased Inquiries

In this section, I show that if **Evidence Externalism** and **Conditionalization** are true, then, given plausible assumptions about our sources of information about the external world, it is possible to set up an inquiry that is rationally biased in favour of a proposition. To start us off, I'll introduce a couple of useful concepts.

1.1 Inquiries, Priors, Plans, Bias

The first is the notion of an inquiry. An inquiry is an evidence-gathering event that takes an agent from an initial information state to a number of new information states. In this essay, I

---

least on an interpretation given by Neta and Pritchard [2007], Williamson [2000], and Goldman [2009].

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

will focus solely on cases where an agent is certain before her inquiry that she will engage in that inquiry. In such cases, idealizing away some complications, we can represent the relevant inquiry using two elements. The first is a finite possibility space W containing worlds that are compatible with the agent's evidence before the inquiry. The second is an evidence function E that maps each world w in W to a proposition, which captures the strongest evidence the agent gains in w as a result of her inquiry. I'll represent any such inquiry using a structure $\langle W, E \rangle$.

The second useful concept is that of a prior credence function, which is just a probability function defined over propositions, which are sets of worlds in W. This reflects the credences that an agent (who is minimally rational) has in various propositions before the inquiry.

The third useful concept is that of an updating plan. An updating plan tells the relevant agent how to update her credences in response to the evidence she receives as a result of her inquiry. For any inquiry $\langle W, E \rangle$, an updating plan may be thought of as a function R from worlds in W to credence functions such that for any two worlds in W, if the agent gains the same evidence in the two worlds, then R recommends the same credence function in the two worlds.[4] **Conditionalization** gives us such an updating plan for any inquiry $\langle W, E \rangle$ and any prior credence function p defined on subsets of W: if R is a conditionalizing plan based on p, then, for any w in W, $R(w) = p(.|E(w))$ (provided $p(E(w)) > 0$).

An updating plan for any inquiry is biased in favour of a proposition just in case it is guaranteed, by the agent's own lights, to raise her credence in that proposition. In other words, for any inquiry $\langle W, E \rangle$ and any rational prior credence function p, an updating plan R is biased in favour of a proposition H if and only if, for any w in W, if c is the posterior credence function that R recommends relative to w, then the agent's posterior credence $c(H)$ in H is greater than her prior credence $p(H)$ in the same proposition.[5] We will call an inquiry rationally biased if, in that inquiry, the agent updates her credences using a biased plan that is

---

[4]For a similar conception of updating plans, see Greaves and Wallace [2006], Schoenfield [2017], and Das [2019].

[5]This notion of bias is different from Salow's [2018] notion. On Salow's view, bias isn't a property of updating plans, but rather of inquiries themselves: for him, an inquiry is biased in favour of a proposition just in case the expected current evidential support for that proposition is lower than the expected posterior evidential support for that proposition.

epistemically rational for her to comply with.

We can now explain how (given some plausible assumptions) **Evidence Externalism** and **Conditionalization** together make rationally biased inquiries possible.

1.2   Externalism, Factivity and Negative Introspection

Let's start by noticing a consequence of **Evidence Externalism**.

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 can entail non-trivial propositions not only about the agent's non-factive mental states (for example, her phenomenal states) but also about the external world. Typically, evidence externalists take factive mental states (like my seeing that there's a hand before me) to be sources of conclusive evidence about states of the external world.

As a result of these commitments, the evidence externalist should reject one of the following two theses:.

> **Factivity**. Necessarily, if an agent's evidence entails P, then P is true.
>
> **Negative Introspection**. Necessarily, if an agent's evidence doesn't entail P, then her evidence entails that it doesn't entail P.

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's white and lit up with trick red lighting that will make any surface look red. If **Factivity** is true, then, in such a scenario, my evidence won't entail that the wall is red. But it may remain compatible with my evidence that I'm seeing that the wall is red. For I might have no idea that the lighting conditions are abnormal. If seeing that the wall is red suffices for me to have evidence that the wall is red, then it will be compatible with my evidence that my evidence entails that the wall is red. So, **Negative Introspection** will be false: even though my evidence doesn't entail that the wall is red, it doesn't entail that it doesn't entail this.

More generally, the argument is this. Take any situation where, from some source of information, an agent gains evidence that entails a non-trivial proposition P about the external

world. Now, we can create a phenomenally indistinguishable situation in which *P* is false, but the agent gains the same information from the same source of information without having any clue that *P* is false.[6] In such a situation, the agent won't be able to rule out the possibility that her evidence entails *P*. So, if **Factivity** is true, then the agent's evidence in such a situation won't entail that it doesn't entail *P*, even though it doesn't entail *P*. Thus, **Negative Introspection** will be false. With respect to cases of this sort, therefore, the evidence externalist must either reject **Factivity** or **Negative Introspection**.

1.3 The Possibility of Rational Bias

This creates the possibility of rationally biased inquiries.

Suppose **Factivity** is false, and I know this. Consider:

> **Red Wall**. I'm about to enter a room and look at a wall. I am now rationally 0.99 confident that the wall is red and the lighting conditions are normal, but assign a credence of 0.01 to the possibility that it might be white but lit up with red light. Suppose I am rationally sure that if the wall is red and the lighting conditions are normal, then I will see that it is red, so my evidence will entail that the wall is red. Moreover, I am also rationally certain that if the wall is white but lit up with red light, it will appear red to me.

If we allow **Factivity** to be false here, then my evidence in the white wall scenario can entail that the wall is red.

---

[6]This assumption may be questioned by Naïve Realists, like Martin [2004] and Fish [2009], who think that cases of veridical perception have a different phenomenal character from cases involving non-veridical perception. But note two things. First, this position has some problematic consequences: it makes it difficult for the Naïve Realist to come up with a positive or negative characterization of bad cases of perception, such as total hallucinations. For discussions of this problem, see Siegel [2008] and Logue [2012]. Second, even Naïve Realists think that when an agent is in a bad case of perception, there is some good case of perception from which the agent can't epistemically distinguish her situation. If that is right, our argument will still go through.

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 gain is that the wall is red. Call this proposition *Red*, the singleton set containing the world r. So, $E(r) = E(w) = Red$. We can depict this structure with the following diagram (where there is a path from a node A to a node B if and only if the world represented by B is compatible with the agent's evidence in the world represented by A):
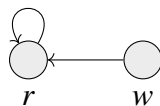

Figure 1: A Failure of **Factivity** in **Red Wall**

If p is my rational prior credence function before entering the room, my prior credence in *Red* will be $p(Red) = 0.99$. Let **Conditionalization** be true. If I update by conditionalizing on *Red*, then my posterior credence in *Red* should be $p(Red|Red) = 1$. So, my inquiry will be rationally biased in favour of *Red*.

If **Factivity** is true but **Negative Introspection** is false, then I can avoid rationally biasing my inquiry in favour of *Red* in **Red Wall**. That situation can be represented as follows.
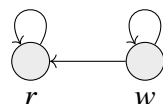

Figure 2: A Failure of **Negative Introspection** in **Red Wall**

This is because, even though my evidence in r will entail *Red*, my evidence in w won't entail it. In fact, $E(w) = \{r, w\}$. So, if I update by conditionalization, my posterior credence in *Red* in w will match my prior credence in *Red*, thus eliminating the possibility of bias.

However, there are other cases where failures of **Negative Introspection** will give rise to rationally biased inquiries (given **Conditionalization**). Take:

> **Red and Sandalwood Wall**. I'm about to enter a room and look at and smell a
> wall. I am now rationally 0.99 confident that the wall is red and made of
> sandalwood. I am rationally sure that if the wall is red, then I'll learn by looking
> that the wall is red, and that if the wall is made of sandalwood, then I'll learn by

smelling that the wall is made of sandalwood. But I rationally assign a credence of 0.005 to the possibility that the wall is white but lit up with red light, and a credence of 0.005 to the possibility that the wall is made of ordinary wood but smeared with sandalwood perfume. Moreover, I have conclusive evidence that the wall won't be both white and made of ordinary wood.

What should happen here?

In this case, I get false information in two possibilities: in the possibility where the wall is made of sandalwood but white, I get the false information that it's red, and in the possibility where the wall is red but made of ordinary wood, I get the false information that it's made of sandalwood. Suppose **Factivity** is false in this case, and I know it. We can represent this inquiry using another simple structure $\langle W, E \rangle$. Here, W contains three worlds: (i) rs (the world where the wall is red and made of sandalwood), (ii) ws (the world where the wall is white and made of sandalwood), and (iii) ro (the world where the wall is red and made of ordinary wood). In rs, the strongest evidence I get is that the wall is both red and made of sandalwood. Call this proposition RS, which is just the singleton set containing rs. So, $E(rs) = RS$. In ws, I learn that the wall is made of sandalwood. If my evidence also entails the false information that the wall is red, then the strongest evidence I get is that the wall is red and made of sandalwood. So, $E(ws) = RS$. Similarly, in ro, not only do I learn that the wall is red, but my evidence may also entail the false information that it's made of sandalwood. So, the strongest evidence I may get is that the wall is red and sandalwood. So, $E(ro) = RS$. We can depict this structure as follows.
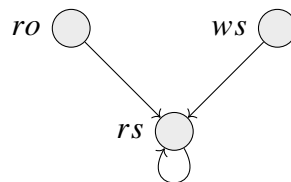

Figure 3: A Failure of **Factivity** in **Red and Sandalwood Wall**

Here, if I update by conditionalization, my credence in RS will rationally increase to 1 no matter which world I am in. Thus, my inquiry will be biased in this case. (Note that a similar result will hold even if we allow **Factivity** to fail in just one of the worlds other than rs.)

Suppose, then, that **Factivity** is true, but **Negative Introspection** is false here. Suppose, also, that I know this. If **Factivity** is true, then, in *ws* where the wall is white and made of sandalwood, my future evidence will entail that the wall is made of sandalwood, but won't entail that it is red. However, since the wall will look red to me, it will remain compatible with my evidence that my evidence entails that the wall is red. Similarly, in *ro* where the wall is red but made of ordinary wood, my future evidence will entail that the wall is red, but won't entail that it is made of sandalwood. Since the wall will smell as if it's made of sandalwood, it will remain compatible with my evidence that I have learnt by smelling that the wall is made of sandalwood, and therefore, that my evidence entails that the wall is made of sandalwood. Thus, **Negative Introspection** will fail in these two scenarios.

Let's say that happens. We can represent this inquiry using the structure $\langle W, E \rangle$. Here, as before, *W* contains three worlds *rs*, *ws*, and *ro*. In *rs*, the strongest evidence I get is *RS*. So, $E(rs) = RS$. In *ws*, the strongest evidence I get is that the wall is made of sandalwood. So, let *WS* be the proposition that the wall is white and made of sandalwood; this is just the singleton set containing *ws*. So, $E(ws) = RS \cup WS$. In *ro*, the strongest evidence I get is that the wall is red. Let *RO* be the proposition that the wall is red and made of ordinary wood; this is just the singleton set containing *ro*. So, $E(ro) = RS \cup RO$. We can depict this structure as follows.
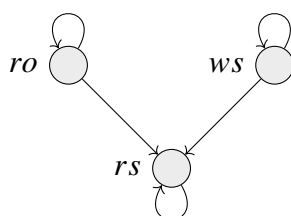

Figure 4: A Failure of **Negative Introspection** in **Red and Sandalwood Wall**

Let *p* be my rational prior credence function before I enter the room, such that $p(RS) = 0.99$ and $p(RO) = p(WS) = 0.005$.

Suppose **Conditionalization** is true. In the scenario where the wall is both red and made of sandalwood, if I update by conditionalization, my posterior credence in *RS* will be $p(RS|RS) = 1$. In the scenario where the wall is white but made of sandalwood, if I update by conditionalization, my posterior credence in *RS* will be $p(RS|RS \cup WS) \approx 0.995$. Finally in the scenario where the wall is red but made of ordinary wood, if I update by

9

conditionalization, my posterior credence in RS will be $p(RS|RS \cup RO) \approx 0.995$. Thus, no matter which scenario I am in, my posterior credence in RS will rise. My inquiry, once again, will be rationally biased.

1.4 An Explanation

These examples suggest that if **Evidence Externalism** and **Conditionalization** are true, then (given plausible assumptions about our sources of information about the external world) it is possible for an agent like us to rationally bias her inquiry. We can make this idea more precise.

Start with a principle that an evidence externalist needn't reject.

> **Positive Introspection**. Necessarily, if an agent's evidence entails P, then her evidence entails that it entails P.

Both **Red Wall** and **Red and Sandalwood Wall** preserve **Positive Introspection**. Even if the evidence externalist rejects this constraint on independent grounds, there's no reason to think it will always fail: even deniers of **Positive Introspection** agree that we often have positive introspective access to our own evidence.[7]

Factivity, **Positive Introspection**, and **Negative Introspection** correspond to three properties of inquiries respectively: reflexivity, transitivity, and euclideanness. **Factivity** is captured by reflexivity: an inquiry $\langle W, E \rangle$ is reflexive just in case, for any world w in W, w is in $E(w)$. **Positive Introspection** corresponds to transitivity: an inquiry $\langle W, E \rangle$ is transitive just

---

[7]Evidence externalists 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] 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. Ahmed and Salow [forthcoming] explore the consequences of failures of **Positive Introspection** for **Value of Information**. 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].

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, **Negative Introspection** is captured by euclideanness: 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 $E(w_2)$. An inquiry is partitional if and only if it has all these three properties. In such inquiries, the evidence function imposes a partition over the possibility space, so that each cell of the partition contains all and only those worlds where the agent's posterior evidence is that cell of that partition. With these properties in mind, we can explain why **Evidence Externalism**, when combined with **Conditionalization**, gives rise to rationally biased inquiries (under plausible assumptions).

Suppose you're an evidence externalist, but you want to reject **Factivity** instead of **Negative Introspection** with respect to cases like **Red Wall**. Then, we can set up a rationally biased inquiry, where the agent has both positive and negative introspective access to her posterior evidence but her posterior evidence entails falsehoods. In many cases like **Red Wall**, 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 that 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 1.** For any serial, transitive and euclidean inquiry $\langle W, E \rangle$, the following two claims are equivalent:
>
> - $\langle W, E \rangle$ is reflexive.
>
> - There exists no regular prior credence function p such that any conditionalizing plan R based on p is biased in favour of some proposition H.[8]

The import: if **Conditionalization** is true, then, for any serial inquiry (conducted by an agent who has regular priors) that satisfies both **Positive** and **Negative Introspection**, satisfying

---

[8]A regular probability function p is such that, for any $w \in W$, $p(\{w\}) > 0$. All proofs are given in the appendix.

Factivity is necessary and sufficient for blocking the possibility that it is rationally biased.

Suppose now that you're an evidence externalist, but you want to preserve **Factivity** and reject **Negative Introspection**. To show how this creates the scope for rationally biased inquiries, we can introduce another constraint on inquiries: divergence. An inquiry $\langle W, E \rangle$ is divergent just in case, for any two distinct worlds w_1 and w_2 in W, if there is a world w_3 that is in both $E(w_1)$ and $E(w_2)$, then there is some world w_4 such that w_2 is in $E(w_4)$ and w_3 is in $E(w_4)$.

This is a bit dense, so let me explain. Suppose $\langle W, E \rangle$ is an inquiry, where W contains at least three worlds w_1, w_2, w_3, such that w_1 is in both $E(w_2)$ and $E(w_3)$. We can depict this as follows:
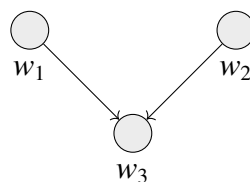

Figure 5

In order to make the inquiry divergent, then we can introduce another world w_4, which is related to w_1 and w_2 in the following way.
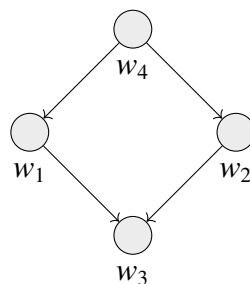
Figure 6

In cases like **Red and Sandalwood Wall**, the inquiry isn't divergent. The world *rs* (where the wall is both red and made of sandalwood) is compatible with my evidence in both *ro* and *ws* where the wall either isn't red or isn't made of sandalwood. But there is no further world where my evidence contains both *ws* and *ro*.[9]

---

[9]However, if a world *wo* (where the wall is both white and made of ordinary wood) were compatible with my evidence before the inquiry, then, in that world, I wouldn't able to rule out

What's the connection between failures of **Negative Introspection** and failures of divergence? Holding **Factivity** fixed, whatever leads to a failure of **Negative Introspection** in **Red Wall** is also responsible for the failure of divergence in **Red and Sandalwood Wall**.[10] Why? In **Red Wall**, when **Factivity** is true, **Negative Introspection** fails because there is a bad case of perception (the world where the wall is white but lit up with red light), where my vision provides me false information without giving me a clue that this has happened. As a result, my evidence doesn't entail the proposition that the wall is red, but also doesn't entail that it doesn't entail it. In **Red and Sandalwood Wall**, divergence fails due to two distinct bad cases of perception: in *ro* and in *rs*, my vision and smell provide me false information, without giving me a clue that this happened. So, **Negative Introspection** fails in these worlds: even though my posterior evidence doesn't entail that the wall is red and made of sandalwood, it remains compatible with my evidence that my evidence entails this. Additionally, in *ro*, I rule out *ws*, and in *ws*, I rule out *ro*, and I have prior evidence that my vision and smell cannot malfunction together. As a result, even though there is a non-empty intersection (containing *rs*) between my evidence in *ro* and *ws*, there is no further world where *rs* and *ws* are compatible with my evidence. This makes my inquiry non-divergent. Intuitively, therefore, an evidence externalist, who preserves **Factivity** and rejects **Negative Introspection** in a case like **Red Wall**, has no good reason to reject the possibility of non-divergence in cases like **Red and Sandalwood Wall**.

More generally, cases like **Red and Sandalwood Wall** provide us with a recipe for creating the possibility that I am either in *ro* or *ws* (provided **Factivity** is true). So, the corresponding inquiry would end up being divergent. But this is not how things are in **Red and Sandalwood Wall**: I antecedently rule out the possibility that the wall isn't both white and made of ordinary wood, so the set of worlds that I cannot rule out before my inquiry doesn't contain any world *wo* where the wall is both white and made of ordinary wood.

---

[10]We can show that a reflexive and transitive inquiry can be divergent only when it violates **Negative Introspection**. Suppose an inquiry $\langle W, E \rangle$ is partitional. Then, for any worlds w_1, w_2, w_3, if w_3 is compatible with the agent's evidence both in w_1 and w_2, then, by transitivity and euclideanness, both w_1 and w_2 must also be compatible with the agent's evidence in w_3. So, the inquiry is divergent.

failures of divergence by exploiting failures of **Negative Introspection**. Suppose an agent has a number of sources of information $S_1, S_2, ..., S_n$, such that (a) the agent can't antecedently rule out the possibility that, in a certain inquiry, each S_i can, independently of the others, provide her false information without giving her any clue that this has happened, but (b) she has prior evidence that if any of them malfunction, exactly one of them will. Then, the resulting inquiry will be divergent (provided that **Factivity** is true).

It turns out that it is possible to rationally bias any non-divergent inquiry where the agent's posterior evidence entails only truths and she has positive introspective access to that evidence. We can show:

> **Proposition 2**. For any reflexive and transitive inquiry $\langle W, E \rangle$, the following two claims are equivalent:
>
> - $\langle W, E \rangle$ is divergent.
>
> - There exists no prior regular credence function p such that any conditionalizing plan R based on p is biased in favour of some proposition H.

The import: if **Conditionalization** is true, then, for any inquiry (conducted by an agent with regular priors) that satisfies both **Factivity** and **Positive Introspection**, satisfying divergence is necessary and sufficient for ruling out the possibility that it is rationally biased.

This completes my argument for the claim that an evidence externalist, who accepts **Conditionalization**, must allow for rationally biased inquiries (given certain plausible assumptions about our sources of information about the external world).

2   The Value of Biased Information

In his argument for **Value of Information**, Good [1967] concerned himself with partitional inquiries. This would block failures of **Factivity** and **Negative Introspection**, thereby ruling out the forms of rational bias we saw in the last section. I am interested in the question of what happens to **Value of Information** when an agent's inquiry is rationally biased.

I'll assume that it is instrumentally rational for an agent only to perform acts that maximize expected value relative to her own current (probabilistically coherent) credence function and her own value function (which reflects the degrees to which she desires different outcomes). We can say this a bit more rigorously. Let a decision problem be a triple $\langle W, A, v \rangle$ where W is a finite possibility space, A is a set of available acts, and v is a value function that maps an act-world pair to the value of performing that act in that world. My assumption about instrumental rationality can be stated as follows.

> **Instrumental Rationality**. For any decision problem $\langle W, A, v \rangle$, if an agent adopts a probabilistically coherent credence function c defined over W, then she is permitted by instrumental rationality to perform an act a in A iff there exists no other act b in A such that:
>
> $$\sum_{w \in W} c(w)v(b, w) > \sum_{w \in W} c(w)v(a, w).^{[11]}$$

For any decision problem $\langle W, A, v \rangle$, I'll let $o(c)$ be an optimal act in A, an act that maximizes expected value relative to c and v.

Suppose **Evidence Externalism** and **Conditionalization** are true. There are two possibilities: either **Factivity** fails in cases like **Red Wall**, or **Factivity** is true but **Negative Introspection** fails in cases like **Red and Sandalwood Wall**. Let's consider these

---

[11] I am being sloppy with notation here: $c(w)$ is just shorthand for $c(\{w\})$. I will assume throughout this essay that states of the world don't depend (either epistemically or causally) on the acts that an agent deliberates about. In fact, complications arise when we relax this assumption and accept either evidential or causal decision theory: evidential decision theorists are led to reject **Value of Information** in Newcomb-style cases where states of the world epistemically depend on the relevant acts, while causal decision theorists have to reject it in cases where states of the world causally depend on the relevant acts. See Skyrms [1990] for discussion of a Newcomb-style case where the evidential decision theorist must reject **Value of Information**, and Rabinowicz [2009] and Ahmed [2014], §7.4.1, for an example involving buying an armour, where gathering and using cost-free evidence is suboptimal according to causal decision theory. I am grateful to an anonymous referee for helpful comments here.

possibilities in turn.

2.1   Failures of Factivity

Suppose **Factivity** fails in **Red Wall**. In that scenario, my prior credence in *Red* (that is, the proposition that the wall is red) is 0.99 but should rise to 1 when I enter the room.

   Now, consider two bets B_1 and B_2 with the payoffs given in Table 1.

	Red	*~ Red*
B_1	1	0
B_2	0.995	0.995

Table 1

Suppose I have no option other than accepting one of these bets, and I can accept only one of them. Relative to my prior credences, the expected value of accepting B_1 is 0.99, while the expected value of accepting B_2 is 0.995. So, I am required by instrumental rationality to accept B_2. In contrast, relative to my posterior credences, the expected value of accepting B_1 is 1 while the expected value of accepting B_2 is 0.995. So, I am required by instrumental rationality to accept B_1.

   Suppose I am rationally certain that I am epistemically and instrumentally rational, and that I am required by epistemic and instrumental rationality to update my credences by conditionalization and perform acts that maximize expected value. So, I can be rationally certain that if I were to act according to my prior credences, I would accept B_2, but if I were to act according to my posterior credences, I would accept B_1. Therefore, by lights of my prior credences, the expected value of acting according to my posterior credences is lower than the expected value of acting according to my prior credences. If the bets in question are offered to me before entering room with the option of postponing the decision until after I've entered the room, I am required by instrumental rationality to make the decision now rather than later.

   This is a failure of **Value of Information**. But this should hardly be surprising. In any scenario where an agent who uses regular priors and updates by conditionalization (and is certain of this) assigns a non-zero probability to a possibility where she receives false posterior evidence, it's possible to create a decision problem where **Value of Information**

16

comes out false.[12] So, the interesting question is whether such failures of **Value of Information** can be induced by failures of **Negative Introspection**.

2.2 Failures of Negative Introspection

Suppose **Factivity** is true, but **Negative Introspection** fails in **Red and Sandalwood Wall**. In that scenario, my prior credence in RS (that is, the proposition that the wall is red and made of sandalwood) is is 0.99. When I enter the room, my credence in RS should rise to either 1 or to 0.995 (approximately).

Now, consider two bets B_1 and B_2 with the payoffs given in Table 2.

	RS	$\sim RS$
B_1	1	0
B_2	0.9925	0.9925

Table 2

---

[12]Here's a proof. Suppose $\langle W, E \rangle$ is an inquiry such that, for some world w^* in W, $w^* \notin E(w^*)$. Now, suppose p is a regular prior credence function that the agent adopts before she gathers evidence. Let r_1 and r_2 be two positive real numbers such that $(1 - p(w^*))r_1 < p(w^*)r_2$. (That there will be such positive real numbers is guaranteed by the regularity of p.) Now, we can construct a decision problem $\langle W, A, v \rangle$, such that A just contains two acts a_1 and a_2 such that (i) $v(a_1, w^*) = -r_2$ and, for any w in W other than w^*, $v(a_1, w) = r_1$, and (ii) for any w in W, $v(a_2, w) = 0$. We can show that the expected value of a_1 relative to p is negative, and therefore less than that of a_2. But if the agent is actually in w^*, then, after gathering evidence and updating by conditionalization, she will assign a credence of 0 to w^*. So, by her lights, the expected value of a_1 will be r_1, and therefore will be greater than that of a_2. Thus, when the agent is in w^*, she will be required by instrumental rationality to choose a_1, and, as a result, will lose r_2. If this is right, then, no matter what other act the agent goes for in the other worlds, the expected value of acting in light of her future credences cannot be positive by lights of the agent's prior credence function. Therefore, **Value of Information** will fail here.

Relative to my prior credences, the expected value of accepting B_1 is 0.99 while the expected value of accepting B_2 is 0.9925. So, I am required by instrumental rationality to accept B_2.

Things are different after I enter the room. If the wall is both red and made of sandalwood, my credence in RS after entering the room is 1. Relative to those credences, the expected value of accepting B_1 is 1 while the expected value of accepting B_2 will be 0.9925. So, I am required by instrumental rationality to accept B_1. Similarly, if the wall is either not red or not made of sandalwood, my credence in RS after entering the room is approximately 0.995. Relative to those credences, the expected value of accepting B_1 is approximately 0.995 while the expected value of accepting B_2 is 0.9925. I should accept B_1.

Suppose I am rationally certain that I am epistemically and instrumentally rational, and that I am required by epistemic and instrumental rationality to update my credences by conditionalization and perform acts that maximize expected value. So, I can be rationally certain that if I were to act according to my prior credences, I would accept B_2, but if I were to act according to my posterior credences, I would accept B_1. Here, relative to my prior credences, the expected value of acting according to my posterior credences is lower than the expected value of acting according to my prior credences. So, if the bets are offered to me before entering room with the option of postponing the decision until after I've entered the room, I am required by instrumental rationality not to postpone my decision.

The upshot: in these scenarios, it is instrumentally irrational for me to gather more evidence and use it for making decisions even if the evidence is available to me for gathering and use at a negligilible cost. Thus, **Value of Information** is false.

2.3   A Diagnosis

In these examples, it's the biased updating plan that leads to failures of **Value of Information**. We can prove:

> **Proposition 3**. For any inquiry $\langle W, E \rangle$ and any prior credence function p defined
> on subsets of W, suppose R is an updating plan that is biased in favour of a
> proposition H. Then, there exists a decision problem $\langle W, A, v \rangle$ such that the
> expected value of acting instrumentally rationally relative to p is greater than the

expected value of acting instrumentally rationally relative to the posterior
credences recommended by R. In other words,

$$\sum_{w \in W} p(w)v(o(p), w) > \sum_{w \in W} p(w)v(o(R(w)), w).$$

This shows that if an agent's updating plan is biased and outputs only probability functions, then we can create a decision problem such that, relative to the agent's prior credences, the expected value of performing an instrumentally rational act in light of her posterior credences is lower than that of performing an instrumentally rational act in light of her prior credences. So, if the agent is certain that she will act in an instrumentally rational manner, then it's instrumentally irrational for her to gather cost-free evidence and use for making her decisions.

In the last section, we saw that if **Evidence Externalism** and **Conditionalization** are true, then, given some plausible assumptions about sources of information, an agent can rationally bias her inquiry. Here, we have seen that, in such rationally biased inquiries, it is instrumentally irrational for certain agents to gather cost-free evidence and use it for making decisions, at least if a certain assumption about instrumental rationality is true. Therefore, if **Evidence Externalism** and **Conditionalization** are true, then, given some plausible assumptions, **Value of Information** is false.

2.4   Connections with Other Similar Results

It's important to highlight why **Propositions 1-3** together reveal a tension that is different from other similar conflicts that have been highlighted in recent discussions of **Value of Information**. (Those who are not interested in the formal features of these results may skip this subsection.)

Ahmed and Salow [forthcoming] notice that non-partitional inquiries and certain forms of risk-aversion can lead to failures of **Value of Information**. But they go on to argue that there is a platitudinous thesis (what they call **Conditionality**) which yields **Value of Information** under ideal conditions (for example, when the agent's inquiry is partitional and the agent isn't risk-averse in certain ways). This thesis, as they convincingly argue, doesn't come out false when the agent's inquiry is non-partitional or when the agent is risk-averse in the relevant

ways. My aim here is different from theirs. While I agree with Ahmed and Salow that there is a kernel of truth in **Value of Information**, but I want to focus on the question of when failures of partitionality lead to failures of **Value of Information**. Ahmed and Salow don't fully address this question.

First, the only non-partitional inquiry that Ahmed and Salow discuss as a counterexample to **Value of Information** involves a failure of **Positive Introspection**. **Positive Introspection** fails in that case due to a controversial margin-for-error principle, discussed by Williamson [2011], which defenders of the KK principle typically reject.[13] One can dismiss the challenge posed by such examples simply by denying such principles. The tension that **Propositions 1-3** bring out amongst **Evidence Externalism**, **Conditionalization**, and **Value of Information** cannot be dismissed in that way: it doesn't depend on the rejection of **Positive Introspection**.

Second, apart from failures of **Positive Introspection**, the only other forms of non-partitionality involve either failures of **Factivity** or failures of **Negative Introspection**. Moreover, we know that **Value of Information** cannot be made to fail for all non-partitional

---

[13]Williamson's [2011] example involves an "irritatingly austere" clock with a completely unmarked dial. Before looking at the clock, the relevant agent's evidence antecedently entails a margin-for-error principle, namely that if the minute hand of the clock is pointing to a number i, then her evidence (after looking at the clock) will entail that it is pointing to a number between $i - 1$ (mod 60) and $i + 1$ (mod 60) (inclusive). This leads to the failures of **Positive Introspection**: when the agent's evidence after looking at the clock entails that the minute hand is between 52 and 54, it remains compatible with the agent's evidence that her evidence doesn't entail this. More importantly, in this case, the agent is also antecedently certain of two facts. First, if the minute hand of the clock is pointing to an odd number, her evidence will support the proposition that it's pointing to an even number. And second, if the minute hand is pointing to an even number, her evidence will support the proposition that it's pointing to an odd number. In such scenarios, since the agent is antecedently certain that her evidence will be misleading with respect to whether the minute hand is pointing to an odd number or an even number, it is instrumentally irrational for the agent to look at the clock and use the evidence she gains to make certain decisions.

inquiries that involve failures of these two constraints.[14] So, the interesting question is when failures of **Factivity** or **Negative Introspection** lead to failures of **Value of Information**. Our results give us a partial answer to this question. **Propositions 1** and **2** show us that in cases where **Positive Introspection** doesn't fail, failures of **Factivity** without failures of **Negative Introspection** or failures of divergence without failures of **Factivity** can lead to rationally biased inquiries. And **Proposition 3** shows us that when an inquiry is rationally biased in favour of a proposition, we can create decision problems relative to which **Value of Information** will fail.

Our results also differ from two less recent results proved by Geanakoplos [1989] (one of which is repeated by Dorst [forthcoming]). The first of these results involves a condition called nestedness. An inquiry $\langle W, E \rangle$ is nested just in case, for any worlds w_1 and w_2, if the intersection of $E(w_1)$ and $E(w_2)$ is non-empty, then either $E(w_1)$ is a subset of $E(w_2)$, or vice-versa. The result in question has two parts. The first part shows that reflexivity, transitivity, and nestedness are together sufficient to preserve **Value of Information** relative to any prior credence function and any decision problem. The second part claims that, for any reflexive, transitive and non-nested inquiry, we can find a prior credence function that will lead to a failure of **Value of Information** relative to some decision problem.

There is an interesting connection between nestedness and divergence: in reflexive inquiries, divergence can fail only if nestedness fails.[15] So, one might worry that the difference between my results and Geanakoplos' result is illusory: if we can show that failures of nestedness in reflexive and transitive inquiries pave the way for failures of **Value of Information**, then failures of divergence in such inquiries should also create the scope for failures of **Value of Information**. But there is a difference between my results and

---

[14]For example, the version of **Red Wall** where **Factivity** holds but **Negative Introspection** fails is a non-partitional inquiry that preserves **Value of Information**.

[15]Suppose, for *reductio*, that a reflexive inquiry $\langle W, E \rangle$ is nested but not divergent. Then, there exist two distinct worlds x, y, such that there is a world z that is compatible with both $E(x)$ and $E(y)$, but there isn't any world w where the agent's evidence contains both x and y. But if the inquiry is nested, then either $E(x) \subseteq E(y)$ or $E(y) \subseteq E(x)$. Suppose $E(x) \subseteq E(y)$. But then, by reflexivity, there exists a world $w = y$ such that $y \in E(w)$ and $x \in E(w)$. Contradiction.

Geanakoplos's nestedness-related result. The latter doesn't tell us whether, for any reflexive, transitive and non-nested inquiry, we can find a regular prior credence function that will lead to a failure of **Value of Information**. Geanakoplos' strategy for proving the second part of his result is to take any reflexive, transitive, and non-nested inquiry and define a non-regular prior credence function on it, so that the inquiry appears to be non-divergent by lights of that prior credence function (though Geanakoplos doesn't explain his strategy in this way and doesn't discuss divergence as a distinct condition on inquiries).[16] This will guarantee (for a conveniently chosen prior credence function) the existence of a decision problem relative to which **Value of Information** will fail. In contrast, **Propositions 2** and **3** guarantee that when divergence fails in reflexive and transitive inquiries, we can find a regular prior credence function relative to which **Value of Information** will fail.

Why does this matter? There is some plausibility to the idea that an agent is required by epistemic rationality to adopt a regular credence function at least when she is distributing her credences over a finite possibility space.[17] If this is right, then Geanakoplos's nested-related

---

[16]It's quite easy to show how this strategy works. Suppose an inquiry $\langle W, E \rangle$ is non-nested. So, there are at least three worlds w_1, w_2, w_3 such that $w_3 \in E(w_1) \cap E(w_2)$ but neither $E(w_2) \subseteq E(w_3)$ nor $E(w_3) \subseteq E(w_2)$. Now, define a prior probability function, such that, for any world w other than these three, $p(w) = 0$. Then, amongst the worlds that the agent antecedently assigns non-zero credence to, there won't be any world w_4, such that $w_1, w_2 \in E(w_4)$. So, by lights of the agent, this inquiry will appear non-divergent. By **Proposition 3**, we can find a decision problem relative to which **Value of Information** will fail.

[17]We can give both evidentialist and pragmatist arguments for this. First, if an agent's evidence doesn't exclude a possibility that a proposition P is true, then the agent's evidence supports P to some positive degree. Any proposition that receives some positive degree of evidential support deserves non-zero credence (at least, in circumstances where assigning non-zero credence doesn't conflict with some other constraint of epistemic rationality). This gives us a partial argument for adopting regular prior credence functions. Second, adopting a non-regular credence function makes an agent exploitable by her own lights. Suppose w is world to which the agent assigns zero credence. We can set a bet that has a negative payoff in

result hasn't given us a clear answer to the question of whether failures of nestedness in reflexive and transitive inquiries lead to a failure of **Value of Information** for agents who use epistemically rational priors. Our results about failures of reflexive and divergence are more informative in that respect, since they give us a recipe for creating failures of **Value of Information** for agents who use regular prior credence functions and engage in non-partitional inquiries.

The second of Geanakoplos' results involves two new conditions: positive balancedness and knowing one's own action. Let a proposition be self-evident relative to an inquiry just in case, for any world compatible with that proposition, the agent's posterior evidence in that world entails that proposition. Using this notion of self-evidentness, we can say what positive balancedness is: an inquiry $\langle W, E \rangle$ is positively balanced if and only if, for any self-evident proposition P, we can find a function λ which assigns to any P-entailing body of future evidence E_i a non-negative value $\lambda(E_i)$ such that, for any w in P, the values assigned by λ to the E_i's that don't eliminate w sum up to 1.[18] The second condition—knowing one's own

$\overline{\hspace{6cm}}$

w and a payoff of 0 everywhere. Now, if betting odds are determined by credences, an agent with the non-regular credence function will accept this bet when it's offered to her for free, since the expected value of the bet is 0. But that means that an agent with a non-regular credence function will often accept bets that involve no possibility of gain, but involve a risk of loss by lights of her own evidence. There is a weak sense in which such an agent is exploitable by her own lights. In contrast, an agent with a regular probability function will turn it down. For defences of regularity (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]. The debate amongst these writers leaves untouched the weak requirement that I am concerned with: namely, when the space of possibilities compatible with an agent's evidence is finite, she shouldn't assign non-zero credence to any of those possibilities.

[18]More precisely, the condition is this. For any inquiry $\langle W, E \rangle$, let a proposition $P \subseteq W$ be self-evident just in case, for any $w \in W$, if $w \in P$, then $E(w) \subseteq P$. Now, for any $X \subseteq W$, suppose $I_X(.)$ is a function that maps a world $w \in W$ to 1 if $w \in X$, and to 0 otherwise. Let $\mathcal{E} = \{E_1, ..., E_k\}$ be the set of the strongest possible pieces of evidence that the agent could get

action—says that for any inquiry $\langle W, E \rangle$, for any prior credence function p and any decision $\langle W, A, v \rangle$, an agent (both before and after gathering and using her evidence) will comply with an action plan such that, in any world w, if the plan recommends an action a from A in w, then, in w, the agent's evidence will entail that the plan recommends that action. Now, this second condition has some plausibility (though it may not be true across the board): often enough, when we act, we know what we are doing, so our own actions are evident to us.

The result that Geanakoplos proves has two parts. Suppose **Conditionalization** is true, and an agent is correctly certain that she is epistemically and instrumentally rational, and also that she satisfies the condition of knowing one's own action. The first part of the result says that any reflexive and positively balanced inquiry conducted by such an agent will preserve **Value of Information**. The second part says that when either reflexivity or positive balancedness fails for such an inquiry, we can find some prior credence function and some decision problem relative to which **Value of Information** will come out false.

Once again, there is an interesting connection between positive balancedness and divergence: we can show that a reflexive and transitive inquiry fails to be divergent only if it isn't positively balanced.[19] So, once again, one might worry that the difference between my

as a result of her inquiry. An inquiry $\langle W, E \rangle$ is positively balanced if and only if for any self-evident proposition P relative to $\langle W, E \rangle$, there exists a function $\lambda : \mathcal{E} \to \mathbb{R}_{\geq 0}$ such that for all $w \in W$, $\sum_{E_i \in \mathcal{E}, E_i \subseteq P} \lambda(E_i).I_{E_i}(w) = I_P(w)$. This definition of positive balancedness, borrowed from Brandenburger, Dekel, and Geanakoplos [1992], p. 185, is a simplified version of the original definition that Geanakoplos [1989] proposes.

[19]Suppose an inquiry $\langle W, E \rangle$ is reflexive, transitive, but not divergent. So, there exist three distinct worlds x, y, z such that $z \in E(x) \cap E(y)$, but there exists no world w such that $x \in E(w)$ and $y \in E(w)$. Let w_x be a world such that $x \in E(w_x)$ but there exists no world w such that $E(w_x) \subset E(w)$. In other words, $E(w_x)$ is one of the weakest evidence propositions that contains x. Let w_y be a world such that $y \in E(w_y)$ but there exists no world w such that $E(w_y) \subset E(w)$. In other words, $E(w_y)$ is one of the weakest evidence propositions that contains y. That there are such worlds is guaranteed by the reflexivity, the transitivity, and the finiteness of the inquiry. Note two facts. By transitivity, (a) $z \in E(w_x)$ and $z \in E(w_y)$, and (b) $E(z) \subset E(w_x)$ and $E(z) \subset E(w_y)$.

results and this one is illusory. But this is not so. Geanakoplos shows that, if **Conditionalization** is true, then, for any reflexive, transitive, and positively unbalanced inquiry, it's possible to find a prior credence function and a decision problem relative to which **Value of Information** fails at least for agents who are certain that they are epistemically and instrumentally rational, and that they satisfy the condition of knowing one's own action. This result, together with the fact that all reflexive, transitive and non-divergent inquiries fail to be positively balanced, isn't sufficient to derive the conclusion that if **Conditionalization** is true, then, for any reflexive, transitive, and non-divergent inquiry, it's possible to find a regular prior credence function and a decision problem relative to which **Value of Information** fails for such agents. This is because there are reflexive, transitive, and positively unbalanced inquiries, for which it is impossible to find any regular prior credence function that will make **Value of**

Suppose, for *reductio*, that the inquiry is positively balanced. Let $\mathcal{E} = \{E_1, ..., E_k\}$ be the set containing the strongest pieces of evidence that the agent could learn as a result of the inquiry. Now, W is a self-evident proposition. So, there exists a function λ with non-negative values such that for any w,

$$I_W(w) = \sum_{E_i \in \mathcal{E}, E_i \subseteq W} I_{E_i}(w)\lambda(E_i)$$

But since the inquiry is reflexive and transitive, for any $E_i \in \mathcal{E}$, $I_{E_i}(w) = 1$ iff $E(w) \subseteq E_i$. So,

$$(1) \quad \sum_{E_i \in \mathcal{E}, E(z) \subseteq E_i} I_{E_i}(z)\lambda(E_i) = I_W(z) = 1$$

$$(2) \quad \sum_{E_i \in \mathcal{E}, E(w_x) \subseteq E_i} I_{E_i}(w_x)\lambda(E_i) = I_W(w_x) = 1$$

$$(3) \quad \sum_{E_i \in \mathcal{E}, E(w_y) \subseteq E_i} I_{E_i}(w_y)\lambda(E_i) = I_W(w_y) = 1.$$

By stipulation, for any $E_i \in \mathcal{E}$, $E(w_x) \subseteq E_i$ iff $E_i = E(w_x)$ and $E(w_y) \subseteq E_i$ iff $E_i = E(w_y)$. So, (2) and (3) imply that $\lambda(E(w_x)) = 1$ and $\lambda(E(w_y)) = 1$. Now, $E(z) \subseteq E(w_x)$ and $E(z) \subseteq E(w_y)$. This implies:

$$(4) \quad \sum_{E_i \in \mathcal{E}, E(z) \subseteq E_i} I_{E_i}(z)\lambda(E_i) \geq 1 + 1.$$

This means (1) is false. So, the inquiry is not positively balanced.

Information fail for an agent who is correctly certain that she is epistemically and instrumentally rational and that she satisfies the condition of knowing one's own action.[20]

Now, as I said earlier, regularity is (plausibly) a requirement of epistemic rationality with respect to finite possibility spaces. So, Geanakoplos hasn't told us when non-partitional inquiries give rise to violations of **Value of Information** for epistemically rational agents who

---

[20]Just consider a variant of **Red and Sandalwood Wall**, where the agent cannot antecedently rule out the possibility that the wall is both white and made of ordinary wood. So, assuming that **Factivity** holds, when she finds herself in that world, she gains no evidence and therefore cannot rule out any of the other possible worlds. So, the resulting inquiry is reflexive, transitive, and non-euclidean inquiry $\langle W, E \rangle$, such that (a) $W = \{rs, ro, ws, wo\}$, where wo is the additional world where the wall is both white and made of ordinary wood, and (b) E is exactly the same as in the original example, except that $E(ws) = W$. This inquiry is not positively balanced: corresponding to the self-evident proposition $P = \{rs, ro, ws\}$, there is no function λ with non-negative values such that for all $w \in W$, the sum of values assigned by λ to the evidence propositions that don't eliminate w is 1.

Now, consider any arbitrary regular prior probability function p and any decision problem $\langle W, A, v \rangle$. Suppose the agent complies with a conditionalizing plan R. Since the agent is instrumentally rational, let f be the action plan that the agent will comply with with in light of her future evidence, such that, for any world w, $f(w) = o(R(w))$. If the agent satisfies Geanakoplos's condition of knowing one's own action, then, in any w, if $f(w) = a$, then in every world $w^* \in E(w)$, $f(w^*) = a$. Now, in wo, the agent's future evidence is the entire possibility space. This means that, for any two $w, w^* \in W$, $f(w) = f(w^*)$. But note also that, in wo, the agent's evidence is the same as it was before her inquiry. So, in wo, her posterior credence function $R(w_4) = p(.|W)$ should be the same as her prior credence function p. Since the agent is correctly certain that she is instrumentally rational throughout, the same act should also be instrumentally rational for her to perform before gathering evidence. So, without any loss of generality, we may assume $o(p) = o(R(w))$, for every w. But then, the expected value of acting in light of her future evidence cannot be lower than that of acting in light of her prior evidence. Thus, **Value of Information** holds.

take their future selves to know their own actions. In this respect, my results are an improvement. For **Propositions 1-3** hold for agents who have regular priors and satisfy the condition of knowing their own actions.

3   Responses

A defender of **Value of Information** cannot easily resolve the conflict with **Evidence Externalism** and **Conditionalization** by rejecting our assumption that expected value maximization is the norm of instrumental rationality. Since expected value maximization turns out to be a special instance of other non-standard rules of instrumental rationality that people have proposed, we could create the same conflict using those non-standard rules.[21] Moreover, certain non-standard norms of instrumental rationality themselves create problems for **Value of Information**.[22] Finding a non-standard norm of instrumental rationality that doesn't itself conflict with **Value of Information** might itself be a difficult challenge to meet. Let us, then, consider some other more promising strategies for solving the problem.

3.1   Impossibility

One strategy will be to say that the cases I've described in §1 aren't really possible. Dorst [forthcoming] adopts this strategy. He argues that the sort of non-divergent inquiry that we see in **Red and Sandalwood Wall** cannot occur. Following Geanakoplos [1989], Dorst accepts the condition called nestedness on inquiries: an inquiry $\langle W, E \rangle$ is nested if and only if, for any two worlds $w, w^* \in W$, if $E(w) \cap E(w^*) \neq \emptyset$, then either $E(w) \subseteq E(w^*)$ or $E(w^*) \subseteq E(w)$. In **Red and Sandalwood Wall**, if **Factivity** remains true but **Negative Introspection** fails, my inquiry isn't nested. For any RO-world where the wall is red and made of ordinary wood, the strongest evidence I gain is $RS \cup RO$, the proposition that the wall is red. For any WS-world where the wall is white and made of sandalwood, the strongest evidence I gain is $RS \cup WS$, the proposition that the wall is made of sandalwood. Even though these two evidence

---

[21]Starmer [2000] is a helpful survey of such non-standard rules of decision-making in descriptive decision theory.

[22]See Buchak [2010] and Campbell-Moore and Salow [forthcoming].

propositions have a non-empty intersection, neither is a subset of the other. In general, failures of divergence of this sort are blocked by nestedness. So, the evidence externalist, who accepts **Factivity** but rejects **Negative Introspection**, could just embrace nestedness as a condition on inquiries.

What should such an externalist say about **Red and Sandalwood Wall**? She could try to preserve nestedness by claiming that my vision and smell cannot malfunction independently of each other. The story will go like this. Suppose **Factivity** is true. Whenever my vision provides me false information that the wall is red, I fail to learn that the wall is red. But then, I also fail to learn that the wall is made of sandalwood. So, in the world ws (where the wall is white and made of sandalwood), my evidence cannot entail $RS \cup WS$. Analogously, whenever I fail to learn by smelling that the wall is made of sandalwood, I fail to learn that the wall is red. So, in the world ro (where the wall is red but made of ordinary wood), my evidence cannot entail $RS \cup RO$. The inquiry can now be depicted as follows.
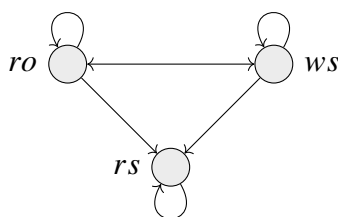

Figure 7: Preserving Nestedness in **Red and Sandalwood Wall**

This will preserve nestedness. For my evidence in both ws and ro will be the same.

However, this proposal fails: it's unclear why the unreliability of my vision should prevent my sense of smell from providing me evidence, and vice-versa. After all, even when one of our senses is defective, we regularly use our other senses to reliably gain information about the external world. Therefore, this proposal looks implausible in light of an externalist picture on which there are multiple independent ways of gaining evidence about the external world.

Suppose we grant that vision and smell can malfunction independently of each other. But we can still make the relevant inquiry nested by making my evidence in ro disjoint from my evidence in ws. This is in fact what Dorst [forthcoming], p. 28, says. What does that imply? We'll have to first enrich the possibility space by adding some new worlds where the wall is both red and sandalwood. So, let the relevant inquiry be $\langle W, E \rangle$, such that

28

$W = \{rs_1, rs_2, ro, ws\}$. Next, let the proposition that the wall is both red and sandalwood be

$RS = \{rs_1, rs_2\}$. On this view, when I see that the wall is red in ro, my evidence comes to

entail that the wall is red, so it rules out ws. But it also rules out one of the RS-worlds like rs_2.

So, $E(ro) = \{rs_1, ro\}$. Analogously, in ws, I not only rule out ro, but also rs_1. So,

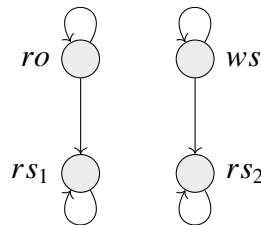$E(ws) = \{rs_2, ws\}$. Thus, the new inquiry can be depicted as follows.

Figure 8: Preserving Nestedness in **Red and Sandalwood Wall**

This inquiry is nested, since my evidence in ro doesn't have a non-empty intersection with my

evidence in ws.

Once again, this is implausible. Suppose I am in ro, where the wall is made of ordinary

wood. And suppose, before actually entering the room, I underwent a simulated version of the

same experience, whereby I learnt exactly how the wall would look and smell to me when I

actually entered the room. When I actually enter the room and see that the wall is red, my

visual experience can give me the evidence that the wall is red. But I do not thereby gain any

evidence that could help me rule out any of the RS-worlds that were previously compatible

with my evidence. In order to do so, I would need more evidence that helps me rule out one of

the worlds where the wall is both red and made of sandalwood. By stipulation, the only

relevant sources of information are vision and smell. In ro, I gain no new evidence from smell.

And while I definitely visually learn something about the colour of the wall when I enter the

room, I don't visually learn anything about the wall other than the fact that it's red. Then, how

can I rule out any of the possibilities where the wall is both red and made of sandalwood?

Thus, the proposal under discussion seems quite arbitrary.[23]

---

[23]Dorst [forthcoming], pp. 27-8, offers an argument for why failures of nestedness are bad.

The argument is driven by an assumption regarding certainties about indicative conditionals:

Epistemic Robustness. Suppose an agent is rationally certain that if $\neg q$, then p. If

The upshot: even though nestedness could help us block failures of divergence in **Red and Sandalwood Wall**, it is unmotivated from the standpoint of the evidence externalist.

3.2 Rejecting Evidence Externalism

Can we save **Value of Information** by rejecting **Evidence Externalism**? Since it is this view that is partly responsible for the tension between **Factivity** and **Negative Introspection** and thereby creates room for biased inquiry, rejecting it may indeed remove the possibility of a rationally biased inquiry. For instance, a Cartesian conception of evidence—on which an agent's evidence only entails propositions about her phenomenal states—might be helpful here. On some versions of the Cartesian view, our evidence consists of facts we know or are in a position to know by introspection about our phenomenal states. If we assume that if we don't know something by introspection, we know or are in a position to know that we don't know it, these versions of the Cartesian view may preserve both **Factivity** and **Negative Introspection**.

However, such a Cartesian view will face at least two problems. First, it's not obvious that **Negative Introspection** cannot be false on such views. Just imagine a scenario where I place my hand under the tap, expecting to be scorched by hot water. In fact, the water is ice-cold.

---

the agent is rationally certain that p, then her epistemic access to the claim that p is more robust than her epistemic access to the claim that q (in othre words, on the supposition that at most one of p or q is true, she'll be rationally certain that p).

This principle seems false. Suppose I learn that p on the basis of some source of evidence, and become rationally certain that p. At this stage, I am not certain that q; by Dorst's own admission, under such circumstances, I can be rationally certain that if $\neg q$, then p. Next, I learn from an independent source of evidence that q. This source of evidence (by my own lights) may be just as reliable as the source of evidence from which I learnt p. Plausibly, if I revise my beliefs monotonically, I can continue to be rationally certain that if $\neg q$, then p. But, contrary to **Epistemic Robustness**, in this situation, my epistemic access to the claim that p isn't more robust than q. If I were to suppose that only one of these two claims—p and q—is true, I needn't continue to be certain that p, given that the two bits of information originate from two equally reliable sources of information by my own lights.

But, for the first few seconds, I misjudge that cold sensation to be a hot one. In such a case, I am not in a position to know by introspection that my sensation is hot. But I am also not in a position to know that I am not in a position to know it. So, even though my evidence might not entail that my sensation is hot, it doesn't entail that it doesn't entail it. So, **Negative Introspection** is false. Thus, even on such a Cartesian view, we can create scenarios of biased inquiry like **Red and Sandalwood Wall** by appealing to two sensations which I might independently misjudge in this way.

Second, if a Cartesian view is combined with **Conditionalization**, we get sceptical consequences.[24] 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 **Conditionalization** is true, the infant can only be rationally confident in the proposition M that there's 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 isn't 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's unclear if we could have *a priori* reasons for discounting contingent sceptical hypotheses.[25] If we can't have such reasons, the Cartesian view will lead to scepticism (when combined with

---

[24]See Neta [2009] for this argument.

[25]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.

Conditionalization).

3.3   Rejecting Conditionalization

The only other strategy is to reject **Conditionalization**.[26] Whatever we might replace **Conditionalization** with, it cannot be a norm of rationality that licenses biased inquiries. For that, by **Proposition 3**, will lead to violations of **Value of Information**.

I don't think this strategy can easily succeed. Suppose **Factivity** fails in **Red Wall**. Since I gain the same evidence whether or not the wall is red, any updating plan will have to recommend the same credence function everywhere. Therefore, I can avoid biasing my inquiry in this scenario only by holding my credences in *Red* (that is, the proposition that the wall is red) fixed. If I were to lower it, my updating plan would be biased in favour of ∼ *Red*; if I were to increase it, my updating plan would be biased in favour of *Red*. Thus, I can only avoid biasing my inquiry by setting my posterior credence in *Red* to 0.99. But, then, there is a sense in which I'll be ignoring my evidence. For consider:

> **The Entailment-Support Principle**. If an agent's evidence entails a proposition
> *P*, then her evidence conclusively supports *P*.[27]

[26]An opponent of **Conditionalization** could give up the propositionalist conception of evidence and, following Jeffrey [1992], could replace it with a probabilistic conception of evidence on which our evidence consists in certain constraints on our posterior credences. She could then appeal to Jeffrey conditionalization as the norm of revising our credences in response to that evidence. However, it's not immediately obvious how this would solve the problem for **Value of Information** posed by rationally biased inquiries. There is nothing in Jeffrey's rule or his conception of evidence, which prevents an agent's credences in a proposition from uniformly increasing when she undergoes a new experience.

[27]The **Entailment-Support Principle** can be resisted. First, in light of cases like **Red Wall**, Neta [2019] denies the claim that an agent is required to be certain of her evidence. It's unclear, however, how well-motivated this move is: in cases like **Red Wall**, the agent has perfect access to her evidence, so it's unclear why she should remain uncertain of what her evidence entails (at least as long as she knows what it entails). Second, an anonymous referee

Plausibly, if my evidence conclusively supports a proposition, then I should be certain in it. If this principle is right, then, even though I can avoid biasing my inquiry by assigning 0.99 to *Red* in this case, I can only do so at the cost of not proportioning my doxastic attitudes to the evidence.

One could argue that the **Entailment-Support Principle** is motivated by the same intuition that motivates **Conditionalization**: namely, that degrees of evidential support can be represented as conditional probabilities on one's evidence. In response, it's worth pointing out a consequence of rejecting the **Entailment-Support Principle**. An updating plan that requires us not to raise our credence in the proposition that the wall is red in **Red Wall** is sceptic-friendly. It recommends 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 adopt such a plan. For, by the same reasoning discussed in §3.2, we'll never justifiably believe anything about the external world!

Suppose **Factivity** is true, but **Negative Introspection** fails in **Red and Sandalwood Wall**. After I enter the room, since I either rule out the possibility that the wall is white or the possibility that it's made of ordinary wood or both, I am ruling out some worlds where *RS* (that is, the proposition that the wall is red and made of sandalwood is false) is false, but am not ruling out any world where it is true. That, intuitively, should count as evidence in favour of *RS*. Why? Consider a principle:

---

has pointed out to me that this principle may also be in tension with the claim all our evidence just is knowledge, since there are many pieces of knowledge of which we may not be required to be certain. Now, some of these cases (such as Radford's [1966] example of the unconfident examinee) are scenarios where the agent lacks perfect access to her knowledge, and therefore aren't analogous to **Red Wall**. Other cases involve cases of inductive or abductive knowledge, where an agent cannot be certain about the conclusion of an inference. Not only are such cases disanalogous to **Red Wall** (which involves perceptual evidence), but they also suggest that we should restrict the status of evidence only to things that we know non-inferentially.

Symmetry of Evidential Support. If the degree of evidential support for P relative to $E \cap Q$ is greater than the degree of evidential support for P relative to E alone, then the degree of evidential support for Q relative to $E \cap P$ is greater than the degree of evidential support for Q relative to E alone.

This principle is plausible. Suppose, right now, my evidence entails that a card has been selected from a random deck of cards but nothing more. If I were to learn now that the card is a five of spades, my evidence would come to conclusively support the claim that the card is black. This also means that if I were to learn now that the card is black, that would give me some evidence that the card is a five of spades.

Apply **Symmetry of Evidential Support** to **Red and Sandalwood Wall**. In **Red and Sandalwood Wall**, before I enter the room, my evidence doesn't conclusively support the proposition that the wall is red. After entering the room, if I were to learn RS, my evidence would come to entail $RS \cup RO$. By the **Entailment-Support Principle**, my evidence would then conlusively support this. By **Symmetry of Evidential Support**, therefore, after entering the room, if I were to learn $RS \cup RO$ (instead of RS), the evidential support for RS should also increase. Once again, here, I can avoid biasing my inquiry only if, in some of these worlds, I don't raise my credence in RS. But that means that I can only avoid biasing my inquiry by ignoring evidence in favour of RS. The result: if the **Entailment-Certainty Principle** and **Symmetry of Evidential Support** are true, then, in this case, the only unbiased updating plans are the ones that require me to ignore the evidence I get.

We could try to reject this argument by rejecting either the **Entailment-Certainty Principle** or **Symmetry of Evidential Support**. But, once again, any principled way of rejecting **Entailment-Certainty Principle** will give rise to sceptical worries just as it did earlier. And **Symmetry of Evidential Support** seems extremely plausible in light of cases like the card example. The result: **Value of Information** cannot be satisfactorily preserved here without at least some intuitive or theoretical costs.[28]

---

[28]In fact, the rule that Schoenfield [2017] calls **Conditionalization*** suffers from these problems. According to this rule, if the strongest evidence that an agent gains between between two times is E and her prior credence function is p, then her posterior credence in

4 Conclusion

Let's take stock. In this essay, I have argued that, given plausible assumptions about instrumental rationality and our sources of information, **Value of Information** conflicts with two other plausible theses, **Evidence Externalism** and **Conditionalization**. I have gone on to claim that every strategy for resolving this conflict involves some cost. So, we cannot easily save **Value of Information**.

Appendix A   Proofs

Proof of Proposition 1. 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 regular prior credence function p such that any conditionalizing plan R based on p is biased in favour of some proposition H. For any proposition $X \subseteq W$, let $[\mathbf{E} = X] = \{w \in W : E(w) = X\}$ be the proposition that the strongest piece of posterior evidence the agent gets is X. Suppose $E_1, E_2, ..., E_k$ 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), $E_i = [\mathbf{E} = E_i]$.[29] Suppose p is a regular probability function and R is a conditionalizing plan

any proposition H should be $p(H|[\mathbf{E} = E])$, where $[\mathbf{E} = E]$ is the proposition that the strongest evidence one has learnt is E. This rule doesn't license biased inquiry. However, it faces the same problems that I mentioned above. In **Red Wall**, if **Factivity** fails, then this would mean that I will update by conditionalizing on $[\mathbf{E} = Red] = Red \cup \sim Red$, which is the entire possibility space. So, my prior credences won't change at all. If the **Entailment-Support Principle** is true, then I will be ignoring evidence. In **Red and Sandalwood Wall**, if **Negative Introspection** fails, in RO-worlds, I will update by conditionalizing on $[\mathbf{E} = RS \cup RO] = RO$, and in WS-worlds, I will update by conditionalizing on $[\mathbf{E} = RS \cup WS] = WS$. So, I will be assigning credence 0 to RS. But, according to the **Entailment-Support Principle** and **Symmetry of Evidential Support**, I gain evidence in favour of RS. Thus, I will be ignoring evidence.

[29]By reflexivity, for any $w \in W$, if $w \in [\mathbf{E} = E_i]$, then $w \in E_i$. Therefore, $[\mathbf{E} = E_i] \subseteq E_i$. By transitivity, for any $w, w^* \in W$, if $w^* \in E(w)$, then $E(w^*) \subseteq E(w)$. By reflexivity and

based on p. So, for any i between 1 and k (inclusive), $R_i(.) = p(.|E_i)$ is the posterior credence function that R recommends in the worlds where the strongest posterior evidence that the agent gains in E_i. Then, by the law of total probability, for any proposition H,

$$p(H) = \sum_{i=1}^{k} p(H|[\mathbf{E} = E_i])p([\mathbf{E} = E_i])$$

$$= \sum_{i=1}^{k} p(H|E_i)p([\mathbf{E} = E_i])$$

$$= \sum_{i=1}^{k} R_i(H)p([\mathbf{E} = E_i])$$

But, then, it cannot be the case that for any i between 1 and k (inclusive), $R_i(H) > p(H)$. Therefore, R isn't biased in favour of any proposition H.

Second, we show that if $\langle W, E \rangle$ is not reflexive, then there exists a regular prior credence function p such that any conditionalizing plan R based on p is biased in favour of some proposition H. 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 isn't. 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 $w \in E(w)$. 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 $E(w^*)$, $p(\sim \{w\}|E(w^*)) = 1$. For any regular probability function p is defined on the subsets of W, $p(\sim \{w\}) < 1$. If an updating plan R is a conditionalizing plan based p, then, for any w^*, $R(w^*) = p(.|E(w^*))$. So, we can conclude that R is biased in favour of $\sim \{w\}$.QED.

Proof of Proposition 2. Suppose an inquiry $\langle W, E \rangle$ is reflexive and transitive.

First, we want to prove that if $\langle W, E \rangle$ is divergent, then there exists no regular prior credence euclideanness, for any $w, w^* \in W$, if $w^* \in E(w)$, then $w \in E(w^*)$. From that, by transitivity, we get: for any $w, w^* \in W$ if $w^* \in E(w)$, $E(w) \subseteq E(w^*)$. So, for any $w, w^* \in W$, if $w^* \in E(w)$, $E(w) = E(w^*)$. Since $w \in E(w)$ by reflexivity, this means that $E_i \subseteq [\mathbf{E} = E_i]$. This entails that $E_i = [\mathbf{E} = E_i]$.

function p such that any conditionalizing plan R based on p is biased in favour of some proposition H. We prove this claim by induction.

- **Base Step**. Suppose the number of worlds in W is 1. In that case, given reflexivity, for any $w \in W$ $E(w) = W$. So, for any regular prior credence function p and any proposition H, $p(H|E(w)) = p(H)$. Thus, there exists no regular prior credence function p such that any conditionalizing plan R based on p is biased in favour of some proposition H.

- **Induction Step.** Suppose the claim we want to prove is true for any inquiry $\langle W, E \rangle$ where the number of worlds in W is at most k. Now, consider an inquiry $\langle W, E \rangle$ where the number of worlds in W is $k + 1$. Suppose, for *reductio*, there exists a prior credence function p such that any conditionalizing plan R based on p is biased in favour of a proposition H. This immediately rules out the possibility that there exists a world $w \in W$ such that $E(w) = W$. For, otherwise, for any regular prior credence function p and any proposition H, $p(H|E(w)) = p(H)$.

  Next, consider any $E(w)$ such that $E(w)$ has the greatest cardinality less than $k + 1$. Since the inquiry is reflexive, there is at least one world in $E(w)$, so the cardinality of $E(w)$ is between 1 and k (inclusive), and the cardinality of $\sim E(w)$ is between 1 and k (inclusive). By transitivity, for any $w^* \in E(w)$, $E(w^*) \subseteq E(w)$, so $E(w^*) \cap \sim E(w) = \emptyset$. From reflexivity, transitivity, and divergence, we get that: for any $w^* \in \sim E(w)$, $E(w) \cap E(w^*) = \emptyset$.[30]

  By the law of total probability, $p(H) = p(H|E(w))p(E(w)) + p(H|\sim E(w))p(\sim E(w))$. Since *ex hypothesi* $p(H) < p(H|E(w))$ and $p(\sim E(w)) > 0$, $p(H|\sim E(w)) < p(H)$. We can construct an inquiry $\langle W^*, E^* \rangle$ where $W^* = \sim E(w)$ and, for any $w^* \in W^*$, $E^*(w^*) = E(w^*)$. Let our regular prior credence function relative to this inquiry be

---

[30]This step might need some explanation. If, for some $w^* \in \sim E(w)$, $E(w) \cap E(w^*) \neq \emptyset$, then there is some world $w^{**} \in E(w)$ such that $w^{**} \in E(w^*)$. By divergence, there is a world z such that $w \in E(z)$ and $w^* \in E(z)$. By transitivity, again, $E(w) \subseteq E(z)$. But then $E(z)$ has a greater cardinality than $E(w)$ since it contains all the worlds in $E(w)$ as well as w^* which isn't in $E(w)$. This contradicts our stipulation.

$p^*(.) = p(.| \sim E(w))$. So, for any $w^* \in W^*$,

$p^*(H \cap W^*) = p(H \cap W^*| \sim E(w)) < p(H) < p(H \cap W^*|E^*(w^*)) = p^*(H|E^*(w^*))$. This

means that any conditionalizing plan based on p^* is biased in favour of $H \cap W^*$. Since

the cardinality of $\sim E(w)$ is at most k, this contradicts our hypothesis. So, the claim to

be proved holds for inquiries where the set of worlds has cardinality $k + 1$.

This completes our proof of the claim.

Next, we show that for any reflexive and transitive but non-divergent inquiry $\langle W, E \rangle$, there

exists a prior credence function p such that any conditionalizing plan based on p is biased in

favour of some proposition H. Suppose $\langle W, E \rangle$ is reflexive and transitive, but not divergent. By

non-divergence, there exist three distinct worlds $x, y, z \in W$ such that $z \in E(x) \cap E(y)$, but there

exists no world w such that $x \in E(w)$ and $y \in E(w)$. Let w_x be a world such that $x \in E(w_x)$ but

there exists no world w such that $E(w_x) \subset E(w)$. In other words, $E(w_x)$ is one of the weakest

evidence propositions that contains x. Let w_y be a world such that $y \in E(w_y)$ but there exists

no world w such that $E(w_y) \subset E(w)$. In other words, $E(w_y)$ is one of the weakest evidence

propositions that contains y. That there are such worlds is guaranteed by reflexivity and

transitivity, and the finiteness of W.[31] Importantly, neither $w_y \in E(w_x)$ nor $w_x \in E(w_y)$.

Now, we can construct a proposition $H = \{w_x, w_y\}$ and a regular prior credence function p

such that (a) $p(w_x) = p(w_y)$ and (b) $p(E(w_x) \cap E(w_y)) > p(\sim (E(w_x) \cap E(w_y)))$. For any

$w \in W$, either $H \cap E(w) = \emptyset$, or not. If the former possibility is true, then

$p(\sim H|E(w)) = 1 < p(\sim H)$. If the latter possibility is true, then either $w_x \in E(w)$ or

$w_y \in E(w)$. By transitivity, either $E(w_x) \subseteq E(w)$ or $E(w_y) \subseteq E(w)$. Since there is no world w

such that $E(w_x)$ or $E(w_y)$ is a proper subset of $E(w)$, this means $E(w_x) = E(w)$ or

$E(w_y) = E(w)$. Suppose that's the case. By (a) and (b), since $p(H \cap E(w)) = p(H \cap \sim E(w))$

_____

[31]By the finiteness of W and reflexivity, for any world w, there is only a finite non-zero

number of worlds w^* such that $w \in E(w^*)$. By transitivity, $E(w) \subseteq E(w^*)$. Now, either some of

these worlds w^* are such that $E(w) \subset E(w^*)$ or there are no such worlds. If there are no such

worlds, $E(w)$ is the weakest body of evidence that contains w. If there are some such worlds,

then we repeat the process again for each such w^*. The finiteness of W guarantees that there is

some world where the agent's evidence is the weakest body of evidence that contains w.

and $p(E(w)) > p(\sim E(w))$, $p(H|E(w)) < p(H|\sim E(w))$. This implies $p(\sim H|E(w)) > p(\sim H)$. So, any conditionalizing plan R based on p is biased in favour of $\sim H$. QED.

Proof of Proposition 3. For any inquiry $\langle W, E \rangle$ and any prior credence function p defined on subsets of W, let R be an updating plan that is biased in favour of a proposition $H \subseteq W$ and only outputs probability functions. Suppose $C = \{x : (\exists w \in W)(R(w) = c \& c(H) = x)\}$ is the possible posterior credences in H that R recommends, and let $c_{min}(H)$ be the lowest of these credences. So, $c_{min}(H) > p(H)$.

Now, consider two acts a and b with the following payoffs:

	H	$\sim H$
a	1	0
b	$\dfrac{p(H) + c_{min}(H)}{2}$	$\dfrac{p(H) + c_{min}(H)}{2}$

Table 3

Let $\langle W, A, v \rangle$ be a decision problem such that (i) $A = \{a, b\}$, (ii) for any H-world $w \in W$, $v(a, w) = 1$ and for any $\sim H$-world $w \in W$, $v(a, w) = 0$, and (iii) for any $w \in W$, $v(b, w) = \dfrac{p(H) + c_{min}(H)}{2}$. Since the inquiry is biased, for any real number $x \in C$, x is greater than the agent's prior credence in H, $p(H)$. So, $\dfrac{p(H) + c_{min}(H)}{2} > p(H)$.

Now, relative to p, the expected value of a is $p(H)$. In contrast, the expected value of b is $\dfrac{p(H) + c_{min}(H)}{2}$ which is greater than $p(H)$. Therefore, for any $w \in W$, if the agent were to act in light of her prior credences, she would be required by instrumental rationality to perform act b. In other words, $o(p) = b$.

Consider next the posterior credence functions. For any w, let $R(w) = c$. Relative to c, the expected value of a is $c(H)$, which is greater than the expected value of b, $\dfrac{p(H) + c_{min}(H)}{2}$. So, if the agent were to act in light of c, she would be required by instrumental rationality to

perform act *a*. So, for any $w \in W$, $o(R(w)) = a$. This means that:

$$\sum_{w \in W} p(w)v(o(p), w) = \sum_{w \in W} p(w)v(b, w)$$

$$> \sum_{w \in W} p(w)v(a, w)$$

$$= \sum_{w \in W} p(w)v(o(R(w)), w).QED.$$

Acknowledgements

Nilanjan Das

Department of Philosophy

University College London

Gower Street

London WC1E 6BT

nilanjan.das@ucl.ac.uk

References

Ahmed, A. [2014]: *Evidence, Decision and Causality*, Cambridge: Cambridge University Press.

Ahmed, A. and Salow, B. [forthcoming]: 'Don't Look Now', *The British Journal for the Philosophy of Science*.

Brandenburger, A., Dekel, E., and Geanakoplos, J. [1992]: 'Correlated Equilibrium with Generalized Information Structures', *Games and Economic Behavior*, **4(2)**, pp. 182–201.

Briggs, R. A. and Pettigrew, R. [forthcoming]: 'An Accuracy-Dominance Argument for Conditionalization,' *Noûs*.

Buchak, L. [2010]: 'Instrumental Rationality, Epistemic Rationality, and Evidence-Gathering', *Philosophical Perspectives*, **24(1)**, pp. 85–120.

Campbell-Moore, C. and Salow, B. [forthcoming]: 'Avoiding Risk and Avoiding Evidence', *Australasian Journal of Philosophy*.

Das, N. [2019]: 'Accuracy and Ur-Prior Conditionalization', *Review of Symbolic Logic*, **12(1)**, pp. 62-96.

Das, N. and Salow, B. [2018]: 'Transparency and the KK Principle', *Noûs*, **52(1)**, pp. 3–23.

Dorst, K. [forthcoming]: 'Evidence: A Guide for the Uncertain', *Philosophy and Phenomenological Research*.

Easwaran, K. [2013]: 'Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection', *Philosophy of Science*, **80(1)**, pp. 119–142.

Easwaran, K. [2014]: 'Regularity and Hyperreal Credences', *Philosophical Review*, **123(1)**, 1–41.

Fish, W. [2009]: *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.

Geanakoplos, J. [1989]: 'Game Theory without Partitions, and Applications to Speculation and Consensus', Cowles Foundation Discussion Paper No. 914, Yale University.

Goldman, A. [2009], 'Williamson on Knowledge and Evidence', in Greenough, P. and Pritchard, D. (eds), *Williamson on Knowledge*, Oxford: Oxford University Press, pp. 73–92.

Good, I. J [1967]: 'On the Principle of Total Evidence', *British Journal for the Philosophy of Science*, **17(4)**, pp. 319–321.

Good, I. J. [1974]: 'A Little Learning Can Be Dangerous', *British Journal for the Philosophy of Science*, **25(4)**, pp. 340–342.

Greaves, H., Wallace, D. [2006]: 'Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility', *Mind*, **115(459)**, pp. 607–632.

Greco, D. [2014]: 'Could KK Be OK?', *Journal of Philosophy*, **111(4)**, pp. 169–197.

Huttegger, S. M. [2014]: 'Learning Experiences and the Value of Knowledge', *Philosophical Studies*, **171(2)**, 279–288.

Jeffrey, R. C. [1992]: *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Kadane, J. B., Schervish, M., and Seidenfield, T. [2008]: 'Is Ignorance Bliss?', *Journal of Philosophy*, **105(1)**, pp. 5–36.

Lewis, D. K. [1980]: 'A Subjectivist's Guide to Objective Chance' of: Jeffrey, R. C. (ed), *Studies in Inductive Logic and Probability*, Volume II, Berkeley: University of California Press, pp. 263–293 .

Logue, H. [2012]: 'What Should the Naïve Realist Say About Total Hallucinations?', *Philosophical Perspectives*, 26(1), 173–199.

Martin, M. G. F. [2004]: 'The Limits of Self-Awareness', *Philosophical Studies*, **120(1-3)**, pp. 37–89.

McDowell, J. [1982]: 'Criteria, Defeasibility, and Knowledge', *Proceedings of the British Academy*, **68**, pp. 455–79.

McDowell, J. [1995]: 'Knowledge and the Internal', *Philosophy and Phenomenological Research*, **55(4)**, pp. 877–93.

McGee, V. [1994]: 'Learning the Impossible', in Eells, E. and Skyrms, B. (eds), *Probability and Conditionals: Belief Revision and Rational Decision*, Cambridge: Cambridge University Press, pp. 179–199.

Neta, R. [2009]: 'Defeating the Dogma of Defeasibility', Greenough, P. and Pritchard, D. (eds), *Williamson on Knowledge*, Oxford: Oxford University Press, pp. 161–82.

Neta, R. [2019]: 'Disjunctivism and Credence', in Doyle, C., Milburn, J., Pritchard, D. (eds), *New Issues in Epistemological Disjunctivism*, New York: Routledge.

Neta, R. and Pritchard, D. [2007]: 'McDowell and the New Evil Genius', *Philosophy and Phenomenological Research*, **74(2)**, pp. 381–396.

Oddie, G. [1997]: 'Conditionalization, Cogency, and Cognitive Value', *British Journal for the Philosophy of Science*, **48(4)**, pp. 533–541.

Peirce, C. S. [1967]: 'Note on the Theory of the Economy of Research', *Operations Research*, **15(4)**, 643–648.

Rabinowicz, W. [2009]: 'Letters From Long Ago: On Causal Decision Theory and Centered Chances', in Lars-Göran, J. (ed), *Logic, Ethics, and All That Jazz - Essays in Honour of Jordan Howard Sobel*, Uppsala: Uppsala Philosophical Studies.

Radford, C. [1966]: 'Knowledge—By Examples', *Analysis*, **27(1)**, pp. 1–11.

Ramsey, F. P. [1990]: 'Weight or the Value of Knowledge', *The British Journal for the Philosophy of Science*, **41(1)**, pp. 1–4.

Salow, B. [2018]: 'The Externalist's Guide to Fishing for Compliments', *Mind*, **127(507)**, pp. 691–728.

Schoenfield, M. [2017]: 'Conditionalization Does Not Maximize Expected Accuracy', *Mind*, **126(504)**, pp. 1155–1187.

Siegel, S. [2008]: 'The Epistemic Conception of Hallucination', in Haddock, A. and Macpherson, F. (eds), *Disjunctivism: Perception, Action and Knowledge*, Oxford: Oxford University Press, pp. 205–224.

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

Skyrms, B. [1990]: 'The Value of Knowledge', *Minnesota Studies in the Philosophy of Science*, **14**, pp. 245–266.

Stalnaker, R. [2015]: 'Luminosity and the KK Thesis', in Goldberg, S. C. (ed), *Externalism, Self-Knowledge, and Skepticism*, vol. 1. Cambridge: Cambridge University Press, pp. 167–196.

Starmer, C. [2000]: 'Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk', *Journal of Economic Literature*, **38(2)**, pp. 332–382.

Teller, P. [1973]: 'Conditionalization and Observation', *Synthese*, **26(2)**, pp. 218–258.

van Fraassen, B. C. [1999]: 'Conditionalization, a New Argument For', *Topoi*, **18(2)**, pp. 93–96.

White, R. [2006]: 'Problems for Dogmatism', *Philosophical Studies*, **131(3)**, pp. 525–57.

Williams, P. M. [1980]: 'Bayesian Conditionalisation and the Principle of Minimum Information', *British Journal for the Philosophy of Science*, **31(2)**, pp. 131–144.

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)**, pp. 173–180.

Williamson, T. [2011]: 'Improbable Knowing', in Dougherty, T. (ed), *Evidentialism and its Discontents*, Oxford: Oxford University Press.

Wright, C. [2004], 'Warrant for Nothing (and Foundations for Free)?' *Aristotelian Society Supplementary Volume*, **78(1)**, pp. 167–212.