

Permissive Rationality and Sensitivity

BENJAMIN ANDERS LEVINSTEIN

Future of Humanity Institute, University of Oxford

Permissivism about rationality is the view that there is sometimes more than one rational response to a given body of evidence. In this paper I discuss the relationship between permissivism, deference to rationality, and peer disagreement. I begin by arguing that—contrary to popular opinion—permissivism supports at least a moderate version of conciliationism. I then formulate a worry for permissivism. I show that, given a plausible principle of rational deference, permissive rationality seems to become unstable and to collapse into unique rationality. I conclude with a formulation of a way out of this problem on behalf of the permissivist.

1. Introduction

It's often far from obvious how to respond rationally to our evidence: Will Hillary Clinton be elected president in 2016? Will Greece leave the Euro by 2020? Is the GRW interpretation of quantum mechanics correct?

The best-informed pundits, economists, and physicists have seriously different views on these questions, and not just because they have slightly different evidence. Even if we made sure all relevant information was shared, substantial difference of opinion would remain. Such agents simply respond to the same evidence differently.

A central question of the theory of epistemic rationality is whether there is nonetheless a single correct response to any batch of evidence.¹ That is, whether the following is true:

UNIQUE: Given any body of total evidence E and proposition p , there is exactly one maximally rational credence to have in p . More precisely, if $\Pr(p|E)$ and $\Pr'(p|E)$ are both maximally rational, then $\Pr(p|E) = \Pr'(p|E)$ for any proposition p and body of total evidence E .

¹ For recent discussion, see (Ballantyne and Coffman 2011; Douven 2009; Feldman 2007; Horowitz 2014; Kelly 2014; Rosen 2001; Schoenfield 2014; White 2005, 2013; Williamson 2010).

At least at first glance, UNIQUE is implausible.² It seems more likely that completely reasonable people can disagree, and that rationality at least sometimes gives us some leeway. That is:

PERMISSIVE: There is at least one body of total evidence E and proposition p such that multiple credences in p are maximally rational. More precisely, there exist proposition p and probability functions Pr and Pr' such that $\text{Pr}(p|E) \neq \text{Pr}'(p|E)$ but both $\text{Pr}(p|E)$ and $\text{Pr}'(p|E)$ are maximally rational credences to have in p on body of total evidence E .

I'll accept for most of this essay PERMISSIVE as a working hypothesis. I want to explore its connection to a couple of related, heavily discussed questions about the nature of epistemic rationality.

QUESTION 1: How should you respond when you discover another agent you took to be just as rational as you disagrees with you under shared evidence?³

Many writers think that if PERMISSIVE is true, conciliationism is wrong.⁴ If Alice has credence .3 that Hillary will win, and she discovers Bob has credence .25, it seems Alice can retain her opinion even if she's aware that Bob is rational too. When asked why, she can appeal to the fact that .3 is perfectly rational, and .3 just struck her as the most attractive credence to have out of the set of rational credences.

QUESTION 2: How should your views about what's rational be coordinated with your first-order credences?

That is, how should an agent's views *about what's rational to think* about a proposition affect her credence in that proposition? While this question has been fairly well-studied under the assumption of unique rationality, relatively little has been said about how to extend it to a permissivist setting for degrees of belief.⁵

² Serious candidates for such a probability function have been advanced by objective bayesians (e.g., Carnap 1950; Jaynes 2003; Rosenkrantz 1981; Solomonoff 1964a, b; Williamson 2010). For better or worse, none has come close to general acceptance.

³ The literature on disagreement is vast. For starters, see (Christensen 2007; Elga 2007; Gutting 1982; Kelly 2005).

⁴ Glossing over some subtleties, we find this view in (Ballantyne and Coffman 2011, 2012; Douven 2009; Enoch 2010; James 1979; Kelly 2005, 2010, 2014; Schoenfield 2014; White 2013). Christensen (2014) has recently argued—for somewhat different reasons than those found here—that permissivism and conciliationism are compatible.

⁵ See (Christensen 2010; Elga 2013; Lasonen-Aarnio 2014; Ross 2006) for discussion deference principles for credence assuming UNIQUE. See also (Coates 2012; Horowitz 2013; Titelbaum 2014; Weatherson ms).

Below I'll present two arguments. Both take as a starting point that rationality is both permissive and substantive. That is to say, rationality doesn't just impose formal coherence constraints (e.g., probabilistic coherence), but also places substantive constraints on agents that guide how they respond to evidence (e.g., that they have a credence between .3 and .4 that Clinton will win the next election).⁶ Moreover, I'll assume that while multiple credences are sometimes rationally permissible, it's somehow or other *categorically epistemically better* to respond rationally to evidence than to respond irrationally, and rational agents are themselves aware of this fact. They recognize further that, while they personally privilege their own rational credence function, there's no objective epistemic reason to privilege it. The tension between the freedom to favor one rational credence function over the others and the mandate to acknowledge the objective egalitarianism of permissive rationality will prove a tough balancing act.⁷

The first argument, which doesn't assume an answer to the coordination question, is for at least *moderate* conciliationism. If Alice and Bob recognize that the other is rational and that they share the same relevant evidence, they should at least modify their credences in p when they discover they disagree. While this argument, if successful, shows that steadfastness is generally the wrong response to disagreement in these circumstances, it falls short of establishing that Alice and Bob should end up with the same credence in p .

The second argument does two things. First, it provides a restricted answer to QUESTION 2. Second, it shows that from there it follows that once r and r' learn they're both rational credence functions, they must end up in complete agreement. If this argument works, permissivism collapses to uniqueness.

I go on to claim that the first argument is correct, but in spite of its unorthodox conclusion, it's not a threat to permissivism *per se*. The second argument is a threat, since it would only allow for permissivism to hold under rational ignorance. I then sketch what I take to be the permissivist's best hope of escape: There's an important evidential difference between rational priors considered as abstract objects and rational priors realized in a particular agent's cognitive architecture. So, agents should sometimes take their attraction to one rational prior over another as genuine evidence for that prior's accuracy. Although this claim may seem desperate, it's more plausible than it seems at first glance. Its ultimate viability, however, will require a more sustained defense than can be given here.

⁶ Horowitz (2014) calls this sort of view *moderate* permissivism, as it counts some coherent priors as definitively irrational. To my mind, this is the natural kind of permissivism to endorse. Whatever else rationality does, it requires agents with my evidence to have credence $>.1$ that it will rain somewhere on earth in the next decade. Strict subjective bayesianism does not rule these absurdly low credences out.

⁷ For an argument against PERMISSIVE that relies on a similar starting point, see Horowitz (2014).

2. The First Argument: Against Steadfasting

2.1. Set-Up and Notation

We'll focus on two canonical agents Alice and Bob with credence functions α and β respectively. Sometimes, these credence functions will be indexed to a time.

Unless otherwise noted, we'll generally assume that they have the same background evidence, know that they share that background evidence, and know that they're both perpetually fully rational.

For ease, I'll also assume $\alpha(p)$ and $\beta(p)$ are sure to be in some discrete set $X=\{0, x_1, \dots, x_n, 1\}$. That is, Alice and Bob know each will adopt one of only finitely many credences. Nothing will hinge on this restriction, but it will allow us to avoid unnecessary mathematical complexity. Scrupulous readers are invited to replace the sums with integrals *mutatis mutandis*.

For the first argument, we'll consider what happens when Alice and Bob receive information about each other's credences in some proposition p . For instance we might want to know what the value of the following should be:

- (1) $\alpha(p|\beta(p) = x)$
- (2) $\alpha_{t_0}(p|\alpha_{t_1}(p) = y)$.
- (3) $\alpha_{t_0}(p|\alpha_{t_1}(p) = y \ \& \ \beta_{t_1}(p) = x)$

Let's establish a notational convention. Because on the right hand side of the vertical bar we'll be considering information about Alice and Bob's credences in some single proposition p at a given time, we'll abbreviate $c_t(p) = x$ as $C_t = x$. The above expressions will then be shortened to:

- (1') $\alpha(p|B = x)$
- (2') $\alpha_{t_0}(p|A_{t_1} = y)$.
- (3') $\alpha_{t_0}(p|A_{t_1} = y \ \& \ B_{t_1} = x)$

This convention has a purpose in addition to just saving space. Alice and Bob will generally be unaware of the other's credence at the point of interest. So, from Alice's point of view, B_t is a random variable describing Bob's credence at t that could take on a number of different values.

2.2. Steadfasting

Let's first consider the following view:

STEADFAST: Suppose Alice knows:

- (1) She's maximally rational.
- (2) Bob's maximally rational.

- (3) She and Bob share total evidence E .
- (4) Rationality is permissive toward p on E .

Then, when she learns Bob's credence in p , she's not rationally obligated to change her credence in p regardless of what Bob's credence turns out to be. That is, for any x , it's permissible that $\alpha(p|B = x) = \alpha(p)$.

I take STEADFAST to be the received view.⁸ There are various reasons philosophers adhere to STEADFAST, but here are two basic lines of thought. First, if PERMISSIVE is true, there's simply no objective reason for Alice to pick one credence over the other. She knows she's responded optimally to the evidence available and that Bob's credence is based on that very same evidence. So, it seems unnecessary for her to change her views merely because Bob has a different attitude. Second, as William James (1979) pointed out long ago, epistemic agents have to balance two conflicting goals: believing what is true and not believing what is false. In the case of full belief, agents can avoid falsity by suspending judgment, but they thereby sacrifice the potential epistemic good of believing truth. Analogously, in a degree-of-belief model, agents can avoid massive inaccuracy by having credences close to .5, but they thereby sacrifice the potential epistemic good of being highly accurate. Plausibly, rationality permits a range of different attitudes toward epistemic risk. If Alice and Bob have different risk-profiles, then they can rationally maintain disagreement.⁹

To counter this view, we'll first see what steadfasting commits Alice to. We first show that in order to ignore Bob, Alice has to think she is vastly better than he is at tracking the truth. In particular, Bob's credence is truth-tracking only insofar as it tracks Alice's own credence. We go on to argue that this self-favoring is incompatible with taking permissive rationality seriously.

2.3. Independence and Conditional Independence

To get an intuitive sense of what's wrong with STEADFAST, we'll need to review some standard probabilistic notions. We begin with:

INDEPENDENCE: Let p and q be propositions and \Pr be a probability function. p and q are *independent according to \Pr* if $\Pr(p|q) = \Pr(p)$.

The basic idea is that if p and q are independent, q doesn't carry any information about p 's truth. For instance, whether I have saag paneer for dinner tonight carries no information (on any reasonable probability function) as to whether it rains in Ankara tomorrow.

⁸ See *fn.* 4 for references.

⁹ See (Kelly 2014) for more recent view taking a similar line.

The following are equivalent to p and q being independent:

- (1) $\Pr(p|q)=\Pr(p|\neg q)$
- (2) $\Pr(q|p)=\Pr(q)$

So, by (1), iff q is uninformative about p , $\neg q$ is as well. By (2), iff q is uninformative about p , p is uninformative about q . *Informative about* is a symmetric relation.

It's important to emphasize that *independence* is a stochastic and not a causal notion. Consider:

- p : Jason watches the Bucks' game.
- q : The Bucks win.

Whether Jason watches the game has no causal bearing on whether the Bucks win. However, suppose my relevant background information is that Jason only watches a game if the Bucks are playing a weak team. Then p and q aren't stochastically independent on my credence function, since the Bucks tend to be more likely to win on days Jason watches them.

A related notion—and one that will be more important for our purposes—is:

CONDITIONAL INDEPENDENCE: Let p , q , and r be propositions and \Pr be a probability function. p and q are *conditionally independent given proposition r according to \Pr* if $\Pr(p|q,r)=\Pr(p|r)$.

The idea here is that once we suppose r , p and q cease to bear any information about one another (if they ever did in the first place).

Consider:

- p : I'll get heart disease.
- q : My blood test indicates I have high cholesterol.
- r : I actually have low cholesterol.

On a reasonable probability function, $\Pr(p|q)>\Pr(p)$. So, once I see my blood test, I should be worried. However, if I later learn that the test was wrong, I no longer have cause for alarm, since what the blood test says no longer tells me anything about my chances of heart disease. So, $\Pr(p|q,r) = \Pr(p|r)$.

The following are equivalent to p and q being conditionally independent given r according to \Pr :

- (1) $\Pr(p|q,r)=\Pr(p|\neg q,r)$
- (2) $\Pr(q|p,r)=\Pr(q|r)$

By (1), iff q is uninformative about p given r , $\neg q$ is as well. By (2), iff q is uninformative about p given r , p is uninformative about q given r .

2.4. Sensitivity to Truth-Values

We'll now apply these standard notions of independence and conditional independence to better understand how one agent might think another agent's credence is tied to the truth of some proposition in question.

If Pr thinks S 's credence is responsive to whether p in any way that's at all informative, then S meets the following condition:

BASIC SENSITIVITY: An agent S is *basically sensitive* to whether p according to Pr if for some x

$$\Pr(S = x|p) \neq \Pr(S = x|\neg p)$$

BASIC SENSITIVITY is relativized to a probability function and is therefore subjective. Nevertheless, it identifies an important kind of sensitivity. An agent meets this condition by the lights of Pr just in case Pr thinks S 's doxastic behavior will vary in some minimally predictable way or other depending on whether p . So, an agent satisfies BASIC SENSITIVITY if Pr doesn't think her credence is always independent of whether p .

The BASIC SENSITIVITY condition is very weak. An agent is basically sensitive if she's just slightly more likely to have a credence of x when p is true than when p is false. She might even be sensitive in the wrong direction. Suppose, for instance, that she's certain to have a credence of .1 when p is true and certain to have a credence of .9 when p is false. Her doxastic behavior strongly tracks whether p , but in the wrong way. That is, one can be *basically sensitive* without being *accurately sensitive*.

However, this notion is still useful when we're gauging Pr's potential response to learning S 's credence. By Bayes' Theorem, we know that if Pr thinks S is basically sensitive to whether p , then for at least some value of x , $\Pr(p|S=x) \neq \Pr(p)$.¹⁰

For our later discussion, it's worth identifying a contrasting and stronger notion of sensitivity that holds of agents Pr fully trusts. First, recall the notion of an *epistemic expert*:

EXPERT: S is an *expert for Pr with respect to p* if for any x , $\Pr(p|S=x)=x$.

Experts are not only sensitive to whether p but also have "uncorrectable" credence functions given the background information at hand. That is, if S is an expert for Pr, then Pr doesn't think it can improve on S 's credence given the evidence it currently possesses.

¹⁰ Note that because BASIC SENSITIVITY is subjective, it only tracks whether Pr thinks learning S 's credence could be informative about whether p . If, for instance, Pr already knows whether p or already knows S 's credence, then S won't count as basically sensitive according to Pr. Although a more objective notion is interesting for assessing an agent's attempts to track the truth, it isn't of primary relevance for our purposes: we're interested in what rational agents think about one another and how they'll update on each other's credences.

Experts (are taken to) have credences that accurately track whether p . In particular, an easy derivation shows:

EXPERT SENSITIVITY: If S is an expert for Pr with respect to p , then for all x :

$$\frac{\text{Pr}(S = x|p)}{\text{Pr}(S = x|\neg p)} \propto \frac{x}{1-x}$$

The important point is the difference in strength between these two sensitivity concepts. If an agent satisfies **BASIC SENSITIVITY**, her credence is *not guaranteed* to be independent of whether p . If she satisfies **EXPERT SENSITIVITY**, her credence is *guaranteed not* to be independent of p (except for the special case in which her credence turns out to be equal to $\text{Pr}(p)$).

2.5. Conditional Sensitivity

For what follows, we'll also need to establish the concept of conditional insensitivity, or guaranteed conditional independence between credence and truth-value.

Let's begin with an example. Suppose Jessica is a doctor, and her patient comes in with signs of disease S . She orders two tests: a nontreponemal and a treponemal test. The former will be back shortly and will be used to determine the initial treatment. If it comes back positive, Jessica will (rightly) have credence .7 the patient has disease S . If it comes back negative, she'll have credence .3. The second test will take a week to come back, but it will completely override the first test. That is, she'll completely ignore the first test and have credence .99 or .01 that the patient has disease S depending only the results of the second test. So we have:

- s : The patient has disease S .
- Test_2 : The second test comes back positive.
- $J_1=x$: Jessica's credence in s after receiving the results only of the first test is x .

Supposing our generic probability function Pr is up to speed on the set up of the case, it will judge Jessica's credence at t_1 to satisfy **EXPERT SENSITIVITY**. However, $\text{Pr}(s|J_1 = .7, \text{Test}_2) = \text{Pr}(s|J_1 = .3, \text{Test}_2)$, and $\text{Pr}(s|J_1 = .7, \neg\text{Test}_2) = \text{Pr}(s|J_1 = .3, \neg\text{Test}_2)$. That is, Jessica's t_1 -credence is *conditionally independent* of whether s given the results of Test_2 . Re-written, we then have:

- $\text{Pr}(J_1 = x|s, \text{Test}_2) = \text{Pr}(J_1 = x|\neg s, \text{Test}_2)$
- $\text{Pr}(J_1 = x|s, \neg\text{Test}_2) = \text{Pr}(J_1 = x|\neg s, \neg\text{Test}_2)$

Conditional on the results of the second test, Jessica's credence loses any sensitivity to the truth of s . Let's then define:

CONDITIONAL INSENSITIVITY: An agent S is *conditionally insensitive to p given q* according to \Pr if $\Pr(S = x|p,q)=\Pr(S = x|\neg p,q)$ for all x .

In other words, an agent is conditionally insensitive to whether p given q , if q renders her credence conditionally independent of p regardless of what her credence turns out to be. Assuming that an agent is unconditionally sensitive to p , conditional insensitivity is hard to achieve. In the case of Jessica, it wasn't enough that the second test was *better* than the first. The first test even if worse could still contribute some independently valuable information. Consider, for instance, two independently conducted political polls that are meant to assess a candidate's chances of election. The first poll has a sample size of 5,000, whereas the second only has a sample size of 500. If you want to predict the results of the election and are offered the results of only one poll, you should clearly opt for the first. However, you'd strictly prefer to know the results of both to knowing the results of either one individually. The reason is that the second poll, although worse, still has new information about whether the candidate will win.

In the disease case, what matters is that the second test is sensitive to whether s in *every way the first test is* and also contributes additional information. The first test is not only less valuable but also adds nothing over and above the second test. It is essentially just equivalent to a noisy version of the second test.

2.6. Back to Steadfastism

With these resources in hand, I'll now argue that STEADFAST permits agents to think they possess vastly superior truth-tracking powers than their rational interlocutors while simultaneously recognizing they have no evidence for this self-favoring view.

To see why, we'll have to take a step back and evaluate Alice's *updating policy* of remaining steadfast. Toward that end, imagine it's t_0 , and at t_0 , Alice knows:

- (1) She and Bob will always be perfectly rational.
- (2) At t_1 , she and Bob will have the same evidence, which may put them in a PERMISSIVE situation.
- (3) Her future credences in p will be her current credence conditionalized on her new evidence.
- (4) At t_2 , she'll learn what Bob's t_1 -credence in p was.
- (5) At t_3 , she won't revise her credence in p .

At t_0 , she's unaware of:

- (A) Her credence at t_1 ,
- (B) Bob's credence at t_1 ,

(C) What evidence they'll get between t_0 and t_1 .

Nevertheless, she's sure that whatever Bob's credence is, she'll ignore it. According to STEADFAST, that's okay. After all, she's sure that her t_1 -self is rational. If Bob is rational at t_1 , and the evidence determines a uniquely rational credence, then they'll agree. If it allows for a range of credences, then STEADFAST says she doesn't have to revise her credence.

We're also assuming that Alice is sure she'll update by *conditionalization*. Updating by conditionalization is perhaps not required in permissive situations. If an agent holds a rationally permissible credence at t_0 and at t_1 , that could well suffice for making her rational. In a permissive situation, her t_1 -credence could still be in the rational range even if it's not the conditionalized version of her t_0 -credence.

However, all we'll need is that updating by conditionalization is *rationally permissible* for an agent like Alice. Surely, if STEADFAST conflicts with conditionalization, it's the former that should go.

Onto the argument. Since Alice updates by conditionalization, we know she obeys the REFLECTION PRINCIPLE. That is:

$$(1) \alpha_0(p|A_1 = x) = x$$

for any x .

Since Alice knows she'll remain steadfast, we have:

$$(2) \alpha_0(p|A_1 = x, B_1 = y) = x$$

for any x and y .

But Equations (1) and (2) hold only if:

$$(3) \alpha_0(B_1 = y|A_1 = x, p) = \alpha_0(B_1 = y|A_1 = x, \neg p)$$

for all x, y . Or, more readably: $\alpha_0(B_1|A_1, p) = \alpha_0(B_1|A_1, \neg p)$.

Let's take a moment to think about what Equation (3) says. Alice thinks her own future credence renders Bob's credence *conditionally insensitive* to whether p . Even though Bob's perfectly rational—not less rational than she is—she thinks he's only tracking p to the extent that his credence is tracking her credence.

FIGURE 1 provides a visual display of Alice's view of the situation. Alice tracks whether p by tying her credence to the evidence she gets. Her credence will vary only when her evidence varies. So, her credence tracks p by responding to her evidence.

What about Bob? Bob isn't *basing* his credence on Alice's in any causal sense. He just receives the same evidence she does and arrives at his credence independently. Nevertheless, as far as Alice is concerned, his credence is *informative* about p only insofar as his credence co-varies with hers. Conditional on either the evidence or on Alice's own cre-

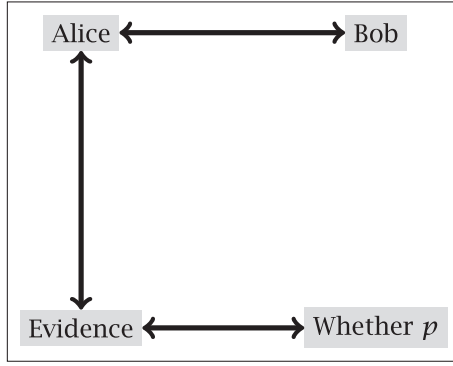


Fig. 1. Information pathways according to Alice

dence, Bob’s credence is conditionally insensitive to whether p . From a stochastic perspective, according to Alice, Bob’s credence is just her credence with noise added.

To cash this idea out, we’ll see which factors determine Alice’s conditional credences that Bob’s t_1 -credence in p is x given that p really is/is not the case. I.e., $\alpha_0(B_1 = x|p)$ and $\alpha_0(B_1 = x|\neg p)$.

We can calculate Alice’s credence that Bob’s credence is x conditional on p by:

- (1) Considering how likely it is that Bob’s credence is x conditional on both p and Alice’s t_1 -credence being x_i
- (2) Discounting that quantity by the probability that Alice’s credence is x_i conditional on p
- (3) Summing over all possible values of Alice’s credence.

And *mutatis mutandis* for $\neg p$. Since both of these quantities will be calculated similarly, we’ll calculate the generic $\alpha_0(B_1 = x|\theta)$, where $\theta=1$ if p and $\theta=0$ if $\neg p$. So,

$$(4) \quad \alpha_0(B_1 = x|\theta) = \sum_{x_i \in X} \alpha_0(B_1 = x|\theta, A_1 = x_i) \cdot \alpha_0(A_1 = x_i|\theta)$$

where X is the set of credences Bob might adopt.

So far, we’ve just adopted the standard operating procedure for calculating the conditional probability of anybody’s credence being x conditional on p or on $\neg p$. But since by Alice’s lights, Bob’s t_1 -credence is conditionally independent of whether p given her own, Equation (4) simplifies to:

$$(5) \quad \alpha_0(B_1 = x|\theta) = \sum_{x_i \in X} \alpha_0(B_1 = x|A_1 = x_i) \cdot \alpha_0(A_1 = x_i|\theta)$$

Equation (5) is just Equation (4) with θ deleted from the first term. What's important here is that Equation (5) brings out Alice's view of the informational connection between whether p and Bob's credence in p .

Alice's t_0 -credence that Bob's credence in p is x conditional on p actually being the case is determined by:

- (1) How Alice thinks her t_1 -credence will track whether p . I.e., the set of credences of the form $\alpha_0(A_1 = x_i | \theta)$.
- (2) How she thinks Bob's credence will track her own t_1 -credences *regardless of whether p* . I.e., $\alpha_0(B_1 = x | A_1 = x_i)$.

So, Alice thinks her future credence will serve as a noisy indicator of p 's truth-value. That is, her future self will have credences that correspond, somewhat roughly, to whether p . Bob's future self will also have a credence function that serves as a noisy indicator of whether p . But there's a second level of noise here. Bob's credence is a noisy indicator of whether p insofar and only insofar as it's a noisy indicator of Alice's own credence.

But the converse can't hold. Alice thinks that her future self will still be sensitive to whether p even conditional on Bob's credence. Otherwise, Alice couldn't both be a steadfast and be sure she'll update by conditionalization.

Thus, we have a strong asymmetry. Alice takes herself to be sensitive to whether p in every way Bob is *and some extra ways too*, but not vice versa. Indeed, even conditional on Bob's credence, Alice's t_1 -credence meets the EXPERT SENSITIVITY condition, whereas Bob is conditionally insensitive to whether p by Alice's lights.

2.7. *Is Alice's Steadfasting Justified?*

It's hard to see how Alice could be justified in her view that she can track the truth in an absolutely superior way to Bob. Both she and Bob are maximally rational and will have the same background evidence. So, Bob's t_1 -credence and her t_1 -credence will be on an objective epistemic par. *Ex hypothesi*, there's no objectively compelling, all epistemic things considered reason to choose one over another. Furthermore, Bob arrived at his credence in the same way Alice did. He considered all the evidence and adopted the credence that he thought was best. Ignoring Bob means that Alice thinks there's nevertheless no independently valuable information to glean from what he thinks.

It also won't do to claim Alice has evidence that she's specially sensitive. First, doing so violates the setup of the case (but more on this later). More importantly, steadfasting is supposed to be an acceptable updating policy for rational agents *generally*. That is, agents are sup-

posed to be rational in steadfasting merely by virtue of the fact that they know they have a rational opinion to start with. Bob, too, would be justified in steadfasting.

Therefore, the steadfaster's position has to be that Alice and Bob are *both* entitled to attribute this special sensitivity power to themselves while admitting they have literally no evidence for their superiority. A rational agent is simply entitled to think that she renders all others conditionally insensitive.

One way to see how odd this conclusion is is to make steadfasters bet on these beliefs. Suppose Alice and Bob share the same background evidence and are rational but steadfast. Let p_1, \dots, p_n be propositions for which Alice and Bob have permissive evidence. Suppose $\alpha(p_i) = a_i > b_i = \beta(p_i)$. Then Bob is willing to sell Alice a bet on p_i for $\frac{1}{2}(a_i + b_i)$ that pays \$1 if p_i and \$0 otherwise. After all, Bob expects the bet to net him:

$$\begin{aligned} & b_i \left(\frac{a_i + b_i}{2} - 1 \right) + (1 - b_i) \cdot \frac{a_i + b_i}{2} \\ &= \frac{a_i + b_i}{2} - b_i \\ &> 0 \end{aligned}$$

Alice is happy to buy the bet, since she expects it to net exactly the same amount:

$$\begin{aligned} & a_i \left(1 - \frac{a_i + b_i}{2} \right) - (1 - a_i) \cdot \frac{a_i + b_i}{2} \\ &= a_i - \frac{a_i + b_i}{2} \\ &= \frac{a_i + b_i}{2} - b_i \end{aligned}$$

It's not hard to concoct situations in which Alice and Bob have radically different expectations about how well they'll do. Suppose all the p_i 's are independent according to both agents and that we've chosen the p_i 's so that $\alpha(p_i) - \beta(p_i)$ is constant. Then as $n \rightarrow \infty$, each will have arbitrarily high credence that he or she will end up with an arbitrarily large amount of money.

Indeed, for n sufficiently large, we can formulate an alternative single bet in which STEADFAST sanctions staking one's life savings on one outcome while admitting that it would have been just as rational to have taken the other side. Consider the following example. Suppose that we've selected a set P_{500} of 500 propositions that both Alice and Bob take to be mutually independent. Let $\alpha(p_i) = .6$ and $\beta(p_i) = .4$ for each $p_i \in P_{500}$. Let h be the

proposition that at least 251 of the 500 propositions in P_{500} are true. Then $\alpha(h) \approx .999996$, while $\beta(h) \approx .000002$.¹¹

Even if Alice and Bob are extremely risk averse, they will still make large bets if their credences are close enough to 0 and 1. Nevertheless, they're fully aware that rationality allows them to hold the opposite view, that only lucky rational agents will end up winning, and that they have no special evidence for thinking they're lucky.

The steadfast may try to claim that Alice is nonetheless entitled to retain her original opinion. But it's unclear how she could reasonably think that steadfasting was an optimal updating policy. That is, it's unclear how she could reasonably expect steadfasting to make her at least as accurate as any alternative policy.

Remember, Alice's goal is not to find some way to avoid updating in light of her interlocutor's credence. Her epistemic goal is to determine whether p . Assuming her credence in p isn't yet 0 or 1, she knows she could be more accurate regardless of how rational she is. She'd be happy to find out that Bob's credence in p provides her with some information about whether p , and she'll try to make any use she can out of it. She should be disappointed if she steadfasts, since she ended up not finding any information of use. The question, then, isn't whether Alice can resist the pressure to update, but whether there is instead pressure not to.

If Alice—despite her best efforts to glean information about whether p from Bob's credence—determines that there's none to be had, it seems she doesn't really take permissive rationality seriously. In some sense or other, she's supposed to recognize that it's better to be rational than not and to recognize she's on a par with all other rational agents. However, hearing from other rational agents who share her evidence, in her view, won't contribute in any way to her inquiry.¹²

¹¹ Since the p_i are all independent, Alice's and Bob's credence in h are given by:

$$\alpha(h) = \sum_{k=251}^{500} \binom{500}{k} .6^k \cdot (1 - .6)^{500-k}$$

and

$$\beta(h) = \sum_{k=251}^{500} \binom{500}{k} .4^k \cdot (1 - .4)^{500-k}$$

¹² Presumably, this point generalizes a bit. If steadfastism is permissible when the same evidence is (known to be) completely shared, then it should still be permissible to mostly ignore agents you think are rational and share nearly all your evidence (as is the case with questions asked in the opening paragraph of this paper).

2.8. Agreeing to Disagree?

So far, I've argued that STEADFAST is wrong. Even if Alice and Bob are rationally permitted to have different credences in p on the same evidence, once they learn the other's view, they should nonetheless revise. But we haven't established *how much* they should revise. Similar considerations to the ones advanced above appear to point to a strong conclusion. Namely, Alice and Bob shouldn't just revise but should end up in full, or near full, agreement at least in ordinary cases. After all, Alice and Bob are objectively on an epistemic par. They both realize there really is no objective reason to go with one credence over another. Therefore, Alice not only lacks evidence that she's superior to Bob at tracking whether p , but she in fact *has* evidence that she isn't, at least not in any significant way. So, once we've admitted that Bob is an independently valuable source of information about whether p and vice versa, we appear to be on the path to admitting that they should take each other to be equally valuable sources of information. That is, they should think they're *equally* sensitive to p 's truth-value.

The Permissivist has enough friction to resist this slide down the slippery slope. There are at least two basic problems with the move from moderate conciliationism to full agreement. The first is with the notion of "equal sensitivity". We were above able to define appropriate notions of sensitivity and conditional sensitivity. But there simply is no obvious generalization of these concepts that would allow us to say when two agents are equally sensitive to whether p . And even if we could define such a notion, it's not obvious that it would require equally sensitive agents to end up in agreement.

The second problem comes with the move from Alice's recognition that she and Bob are on an epistemic *par* to thinking she and Bob are epistemic equals in whatever respect would be necessary for full agreement. They may both be maximally rational, but that doesn't even imply they're *equally* rational, perhaps because they have different epistemic values or standards. Recognizing that Bob is on a par should be enough for Alice not to think that her credence renders his conditionally insensitive to whether p , but the permissivist might still deny that she must end up in agreement with him.

In the next section, I'll present a short argument for total conciliationism: if two credence functions learn they're both rational and share the same evidence, then they must agree. So, by this argument, once the rational credence functions are aware of exactly what is and isn't a rational response, PERMISSIVE collapses into UNIQUE. While the first argument is not necessarily a threat to permissivism, the second argument is.¹³

In the final section, I'll gesture at a potential escape for the permissivist from this collapse argument that runs along the following lines: Alice can

¹³ At least it is a threat to nearly all forms of permissivism in the literature. For a view that accepts a kind of collapse argument, see (Cohen 2013).

learn two important facts when she learns Bob's credence. First, she can gain knowledge about which credences are rational on her background evidence. Since she already knew Bob's credence function was sure to be rational, when she learns $\mathfrak{b}_1(p) = x$, she knows that x is a rational credence to have in p given her evidence. Second, she learns that *Bob* has credence x in p at t_1 . This is empirical information about which credence function is realized in another agent's cognitive architecture, not abstract information about rationality as such.

I'll argue below that if the permissivist ought to take the second fact as epistemically important. Otherwise, she faces a dilemma: either we shouldn't take permissive rationality that seriously, or when the rational credence functions learn about one another, they'll all end up in agreement. The permissivist should maintain that, in many cases, agents should think that the fact that they chose a particular prior can be *evidence in favor of that prior's accuracy*. That is, agents should tend to think of themselves as truth-measurement devices whose readings will beat those of a random rational credence function.

3. The Second Argument

3.1. Deference Principles

To understand the role particular privileged probability functions play in our epistemic lives, philosophers often look to deference principles. A deference principle tells you when you ought to treat a given probability function as an *expert*. For instance, we have the famous PRINCIPAL PRINCIPLE from (Lewis 1980) and the NEW PRINCIPLE from (Hall 1994) and (Thau 1994) that tell us when to defer to chance. As we've already seen, we also have the REFLECTION PRINCIPLE from (van Fraassen 1984) that requires agents to defer to their future time-slices.

Rationality, too, is privileged and sometimes deserves our deference. Elga (2013) defends a principle he dubs NEW RATIONAL REFLECTION to handle cases of unique rationality that behaves analogously to the NEW PRINCIPLE.

If we're going to take permissive rationality seriously and treat it as preferable to irrationality, we need some principle that tells us when to treat rational functions as experts. Below, I'll first propose and defend an extremely weak version of a deference principle to permissive rationality: roughly, there's at least one credence function that should defer to any rational credence function. I show that even this principle leads to collapse. If it's correct, no two rational credence functions can know they're both rational and disagree about anything under shared evidence. I then turn, in §4, to the permissivist's best hope for escape.

3.1.1. The Unique Case

For comparison and simplicity, let's begin with the easier case. Suppose **UNIQUE** is true. There is, then, some single maximally rational function. We'll use $\overset{\circ}{r}$ as a variable to serve as the *description* of that function. That is, $\overset{\circ}{r}$ refers to the rational function, whatever it is.

How should we advise agents to defer to rationality? As a *first pass*, we might suggest the following principle:

FIRST PASS UNIQUE (FPU): If E is \mathfrak{b} 's total evidence, then $\mathfrak{b}(p|\overset{\circ}{r}(p|E) = x) = x$

The idea behind FPU is simple: Supposing you were to find out that on your evidence, the rational credence function $\overset{\circ}{r}$ assigned credence x to p , you ought to assign credence x as well.¹⁴

FPU seems close to right, but it needs a little cleaning up. As Elga (2013) demonstrates, problems arise in cases where the ideal credence function may not know that it's ideal.

Suppose, for instance, that you were looking at the "irritatingly austere" clock in **FIGURE 2** whose minute-hand moves in discrete one-minute jumps.¹⁵ Elga sets up the puzzle thus:

If your eyes are like mine, it won't be clear whether the clock reads 12:17 or some other nearby time. What should you believe about the time that the clock reads?

That, it seems, depends on what the clock really reads. If the clock really reads 12:17, then you should be highly confident that it reads a time near 12:17—99% confident, say, that the time is within a minute of 12:17. But you should be highly uncertain as between 12:16, 12:17, and 12:18. (p. 128)

The idea is that, were the time different, your visual evidence would be different, so what it was rational to believe would vary. If it were 12:45, for instance, it would be far less rational for you to be highly confidence that it's 12:17.

But now we have a problem. Let *12:17ish* refer to the proposition that it's either 12:16, 12:17, or 12:18. Suppose you learn that given your evidence $\overset{\circ}{r}(12:17ish) = .99$. From that information, you can deduce that it's *exactly* 12:17. Here's why: Suppose it were, say, 12:16. Rationality would then assign higher than .01 credence that it's 12:15 and, therefore, lower than .99 credence that it's 12:17ish. So, since rationality actually assigns

¹⁴ Note we could have formulated the principle as:

FPU': If E is \mathfrak{b} 's total evidence, then $\mathfrak{b}(p|\overset{\circ}{r}(p|E) = x, E) = x$

However, since E is \mathfrak{b} 's total evidence, $\mathfrak{b}(\cdot|E) = \mathfrak{b}$. The formulation in the main text thus avoids redundancy.

¹⁵ The original example and term 'irritatingly austere' are due to Timothy Williamson.

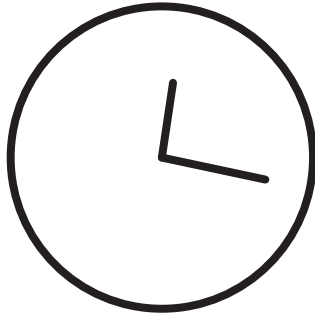


Fig. 2. The Irritatingly Austere Clock

credence .99 that it's 12:17ish, it must not be 12:16. For similar reasons, you can rule out any other time. Therefore, if your credence function is \hat{b} , then $\hat{b}(12:17 | \hat{r}(12:17ish) = .99) = 1$. Since if it's 12:17, it's 12:17ish, we then have $\hat{b}(12:17ish | \hat{r}(12:17ish) = .99) = 1$ *contra* FPU.

What went wrong is that, given the evidence, the ideally rational function assigns credence less than 1 that it's the rational credence function.¹⁶ In other words, if \mathbf{r} is actually the rational credence function, $\mathbf{r}(\mathbf{r} = \hat{\mathbf{r}}) < 1$. So, we need to revise FPU to take this epistemic humility into account:

NEW RATREF: If E is \hat{b} 's total evidence, then $\hat{b}(p | \hat{r}(p|E, \mathbf{r} = \hat{\mathbf{r}}) = x, \mathbf{r} = \hat{\mathbf{r}}) = x$

NEW RATREF brings \mathbf{r} up to speed both on the evidence E and on the fact that it itself is the rational credence function. Since \hat{b} is supposing this to be the case, \mathbf{r} should be making the same supposition.

Is NEW RATREF right? It's at least plausible and in the ballpark. And it will provide us with a foil for the PERMISSIVE case. So, we'll make do with it and turn to the case of interest.

3.1.2. The Permissive Case

We won't even try to suggest anything with the same generality of NEW RATREF for permissive rationality. For our purposes, we'll only need to identify some very minimal principles—much weaker than NEW RATREF—to push the collapse argument through. The principle we come up with will be based on the idea that rationality is at least in some circumstances categorically preferable to irrationality. That is, at least sometimes, one ought to

¹⁶ The problems are similar to problems that arise with self-undermining chance functions. Our new expert principle will be analogous to the NEW PRINCIPLE for chance. Whether the ideally rational function could be unaware that it's rational is controversial. See (Titelbaum 2014).

choose to be rational over being irrational *regardless of which particular function(s) turn out to be rational*.

We'll first restrict attention to probability functions that know they're not entirely rational. If \hat{b} thinks that it might be a rational credence function and then learns that \hat{b} definitely is a rational credence function, it's not at all obvious how \hat{b} should react. We won't try to figure that out.

Now, one might object that appealing to the behavior of irrational credence functions isn't a good guide to understanding the normative force of rationality. More pointedly: why would we think anything important follows from the behavior of an irrational credence function?

I think the answer lies in the role rationality ought to play in regulating our epistemic lives. Rationality should act, in some way or other, as a guide to our doxastic behavior. Even though some agents may be irrational—i.e., not maximally rational—they're not necessarily epistemically hopeless cases. Indeed, they should be able to improve themselves through learning what is, in fact, rational epistemic behavior and somehow imitating it.

Even if irrational agents don't actually respond in the right way to evidence *about* rationality, we can still tell them how they *ought to* respond, i.e., what credal state they should most prefer to be in given the evidence at their disposal. The deference norms we come up with are advice for agents who fall short of rational ideals, even if the agents themselves don't actually always follow the advice. Thus, they appeal not to the actual behavior of irrational credence functions but to the epistemic preferences irrational agents should have.

Let's now move forward and see what a good deference norm might look like. Suppose \hat{b} is certain it does not hold a maximally rational opinion toward some proposition p . It's then told what *some* rationally ideal credence function or other thinks about p . Building on the idea that agents should prefer rationality to irrationality, our first-pass thought is that it ought to defer entirely to that credence function's opinion of p given E . Letting $\mathring{\mathcal{R}}(p, E)$ be the set of rationally permissible credences toward p given total evidence E , we can then capture this idea more formally as follows:

FIRST PASS PERM (FPP): If E is \hat{b} 's total evidence and $\hat{b}(\hat{b}(p) \in \mathring{\mathcal{R}}(p, E)) = 0$, then $\hat{b}(p|x \in \mathring{\mathcal{R}}(p, E)) = x$.

A problem arises immediately. Suppose p_1, p_2 , and p_3 form a partition. That is, they're jointly exclusive and mutually exhaustive. You realize your current opinion toward each of them is not rational. You then learn $\mathbf{r}_1, \mathbf{r}_2$, and \mathbf{r}_3 are all rationally permissible and that $\mathbf{r}_1(p_1|E) = x, \mathbf{r}_2(p_2|E) = y, \mathbf{r}_3(p_3|E) = z$. However, $x+y+z \neq 1$. You then can't defer to each of $\mathbf{r}_1, \mathbf{r}_2$, and \mathbf{r}_3 in the way FPP recommends while maintaining probabilistic coherence.

The problem is that FPP is a *local* expert principle. It tells you to defer on a proposition-by-proposition basis. You don't need to learn what \mathbf{r} thinks about *every* proposition but just need to learn what it thinks about p .

It's possible that there is some correct local principle of deference for permissive rationality. However, we'll concentrate on global principles so we don't have to worry about these interference effects.

So, as a second pass, we'll simply transform FPP into a *global* principle. Let $\overset{\circ}{\mathcal{R}}$ be the set of rational priors, and let $\overset{\circ}{\mathcal{R}}(E) = \{\mathbf{r}(\cdot|E) : \mathbf{r} \in \overset{\circ}{\mathcal{R}}\}$. We arrive at:

SECOND PASS PERM (SPP): If E is \mathfrak{b} 's total evidence and $\mathfrak{b}(\mathfrak{b} \in \overset{\circ}{\mathcal{R}}(E))=0$, then $\mathfrak{b}(\cdot|\mathbf{r} \in \overset{\circ}{\mathcal{R}}) = \mathbf{r}(\cdot|E)$.

SPP is an improvement. It says that if \mathfrak{b} is certain it hasn't responded to its evidence in an entirely rational way, but it learns \mathbf{r} is a rational prior, it should simply feed \mathbf{r} all of its evidence and then defer to it.

But we still don't have an acceptable deference principle. First, as with FPU, we need to correct for any potential self-undermining. That is, we need to let \mathbf{r} know that it's in $\overset{\circ}{\mathcal{R}}$. Second, we need to make sure that the background evidence E doesn't contain any illicit information about what is and is not in $\overset{\circ}{\mathcal{R}}$. We only want to consider cases in which \mathfrak{b} is (1) sure that it itself is not rational, and (2) does not know what other functions are rational.

We then arrive at a reasonable and weak expert principle:

PERM EXPERT: If

- (1) E is \mathfrak{b} 's total evidence
- (2) $\mathfrak{b}(\mathfrak{b} \in \overset{\circ}{\mathcal{R}}(E)) = 0$, and
- (3) For all probability functions \mathbf{r} , $\mathfrak{b}(\mathbf{r} \in \overset{\circ}{\mathcal{R}}) < 1$, then $\mathfrak{b}(\cdot|\mathbf{r} \in \overset{\circ}{\mathcal{R}}) = \mathbf{r}(\cdot|E, \mathbf{r} \in \overset{\circ}{\mathcal{R}})$.

Here's the idea. Suppose you know you're not maximally rational, and you don't know which functions are maximally rational. You then learn that some function \mathbf{r} is rational. PERM EXPERT says after you feed \mathbf{r} all your background evidence E and bring it up to speed on the fact that it's rational, you should just abandon your original credence function \mathfrak{b} and then switch over to it.

I think PERM EXPERT is at least plausible as a permissive expert principle. If you know you're not rational but then learn \mathbf{r} is, you now have three options:

- (1) Stay with some function you're certain is not rational.
- (2) Switch to some function that you think might or might not be rational.
- (3) Switch to a function that you're certain is fully rational.

The last of these looks best, at least at first glance. If you know that as things stand you've responded to evidence in a not-entirely-rational way, then learning that some function \mathbf{r} is rational should at least move you to

revise some of your credences. And, since \mathfrak{r} is the only function you know is rational, switching to \mathfrak{r} appears to be the way to go. Otherwise, you risk remaining irrational when you had the opportunity to become completely rational.

For the collapse argument, we can weaken PERM EXPERT even further. Consider:

WEAK PERM EXPERT (WPE): For some \mathfrak{b} with total evidence E and for all probability functions \mathfrak{r} , $\mathfrak{b}(\cdot|\mathfrak{r} \in \overset{\circ}{\mathcal{R}}) = \mathfrak{r}(\cdot|E, \mathfrak{r} \in \overset{\circ}{\mathcal{R}})$.

WPE says that for *some function or other*, if that function learns another function is rational, it should just switch over to it after feeding it all relevant information. Even if a particular function \mathfrak{b} has reason not to switch to \mathfrak{r} after learning \mathfrak{r} is rational, there is at least one function, according to WPE, that ought to change over to any rational function if given the opportunity.

If there isn't even one probability function that ought to switch to a rational credence function (as WPE would require), then it seems that rationality has nothing special going for it *as a class*. Rationality is not taken to be *categorically* more valuable than irrationality, nor is it *categorically* more conducive to anybody's reasonable epistemic ends than irrationality. No agent, in other words, should think "be rational, however you do it" is good advice. There would be no answer to the question of why we should be rational, since rationality isn't, in general, preferable to irrationality.

Now, WPE may not appeal to all Permissivists, even those who agree rationality imposes substantive constraints. They might think that every irrational credence function has some kind of *epistemic defect*, not that rational credence functions have any special virtue in general (halo notation notwithstanding). According to them, for every irrational credence function, there's some member or other of $\overset{\circ}{\mathcal{R}}$ that is objectively more choiceworthy than it is (i.e., the one or ones that correct the defect). It's not the case, on this view, that *every* element of $\overset{\circ}{\mathcal{R}}$ is more choiceworthy.¹⁷ Indeed, on such views, there will likely be no general norm of deference to rationality. Learning that some function \mathfrak{r} is rational does nothing to make it attractive in general.

However, I think a number of Permissivists will and should balk at this retreat and opt for a meatier conception of rationality. They want to make

¹⁷ For comparison, consider the accuracy dominance argument for probabilism from (Joyce 1998). Given some background assumptions, it can be shown that for any probabilistically incoherent function, there's some probabilistically coherent function that is more accurate than it is at every world. In other words, each incoherent function is *accuracy dominated* by some coherent function. Furthermore, no coherent function is accuracy dominated by any function. So, in one sense, probabilistic incoherence is epistemically defective. However, there's nothing obviously more valuable about coherence *per se*. At the actual world, some incoherent functions will still end up more accurate than some coherent ones.

sense of the idea that rationality as such is somehow special and preferable. It's not simply the case that rational credence functions *lack* a certain kind of epistemic defect, but they have a certain kind of epistemic virtue. Rational agents should afford one another a kind of respect, it seems, that they oughtn't give to irrational agents. We should in general take rational opinion and argument seriously. I'll assume that a principle like WPE will appeal to these permissivists because it recognizes that rationality *as such* is desirable or at least more desirable than at least *one* irrational function.¹⁸

3.2. Collapse

Unfortunately, PERMISSIVE appears to collapse given WPE. Suppose at t_0 , you have credence function \mathfrak{b} and total evidence E , where \mathfrak{b} is such that for any probability function \mathfrak{r} , $\mathfrak{b}(\cdot|\mathfrak{r} \in \mathring{\mathcal{R}}) = \mathfrak{r}(\cdot|E, \mathfrak{r} \in \mathring{\mathcal{R}})$. That is, \mathfrak{b} is a witness to the existential claim in WPE. Consider the following two sequences in which you may receive evidence. The first: at t_1 , you learn $\mathfrak{r} \in \mathring{\mathcal{R}}$, and at t_2 , you learn $\mathfrak{r}' \in \mathring{\mathcal{R}}$. Then, your credence function at t_1 is $\mathfrak{b}(\cdot|\mathfrak{r} \in \mathring{\mathcal{R}}) = \mathfrak{r}(\cdot|\mathfrak{r} \in \mathring{\mathcal{R}}, E)$. At t_2 , your credence function is

$$(6) \quad \mathfrak{b}(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}) = \mathfrak{r}(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}, E)$$

The second: at t'_1 you learn $\mathfrak{r}' \in \mathring{\mathcal{R}}$, and at t'_2 , you learn that $\mathfrak{r} \in \mathring{\mathcal{R}}$. Your credence function at t_1 is $\mathfrak{b}(\cdot|\mathfrak{r}' \in \mathring{\mathcal{R}}) = \mathfrak{r}'(\cdot|\mathfrak{r}' \in \mathring{\mathcal{R}}, E)$. At t'_2 , your credence function is

$$(7) \quad \mathfrak{b}(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}) = \mathfrak{r}'(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}, E)$$

So, $\mathfrak{r}(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}) = \mathfrak{r}'(\cdot|\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}})$ for any $\mathfrak{r}, \mathfrak{r}' \in \mathring{\mathcal{R}}$.

Therefore, as long as something like WPE is right, PERMISSIVE appears to collapse under rational omniscience, at least as long as evidence is commutative.¹⁹ Should all the functions in $\mathring{\mathcal{R}}$ know which functions are in $\mathring{\mathcal{R}}$, only one rational credence function remains. And even when only two rational functions learn of one another, there's a unique rational response. That is, WPE is incompatible with:

¹⁸ I do not mean to rule out the possibility that Permissivists of this sort might have qualms even with a principle as weak as WPE. However, I think it's nonetheless important to show (1) how a simple and *prima facie* attractive principle leads to collapse, and (2) how they might be able to accept it as an abstract principle assuming the considerations appealed to in §5 pan out.

¹⁹ An updating policy treats evidence as commutative if the result of updating on one proposition p and then another q yields the same result as updating first on q and then on p . So, abandoning commutativity means that an agent would sometimes treat the same body of total evidence differently based on the order in which she processed the same information. Note that conditionalization, in particular, is commutative.

DISAGREE: It's sometimes the case that for distinct $r, r' \in \overset{\circ}{\mathcal{R}}$ and some proposition p , we can have $r(p|r, r' \in \overset{\circ}{\mathcal{R}}, E) \neq r'(p|r, r' \in \overset{\circ}{\mathcal{R}}, E)$.

FIGURE 3 displays how the collapse argument works visually. Let r^+ (r'^+) be r (r') fed \mathfrak{b} 's background evidence and the information that it's rational. If \mathfrak{b} learns first that r is rational, by WPE, it will update to state r^+ . If it next learns that r' is rational, it will update again to a new state, which we'll call r^* . Alternatively, if \mathfrak{b} first learns that r' is rational, it will first update to r'^+ . If it next learns that r is rational, it will update to some new state, which by the commutativity of evidence, must be r^* .

4. Sensitivity Over and Above Rationality

Accepting WPE and the collapse result that comes with it isn't strictly speaking incompatible with a permissivist view. After all, multiple responses to evidence could still be permitted so long as rational credence functions don't know about each other. I take it, however, that most permissivists want rationality to be permissive even when rational functions are aware of the alternative rational choices.²⁰ So, permissivists need some way to take rationality seriously while avoiding the collapse results above.

The best path has already been suggested. Some agents, when thinking of themselves as truth-measurement devices, ought to take themselves to be sensitive in ways a randomly selected rational credence function isn't. At least sometimes, Alice should take the fact that she found credence function α more attractive than credence function α' as *new evidence for the superiority of α to α'* .

Before deciding whether we think this view is plausible, let's establish that the permissivist should herself already think it's true. Suppose, *pace* the second argument, that permissive rationality doesn't collapse. That is, suppose there are multiple credence functions that know that they're each perfectly rational, yet nonetheless disagree. If Alice has chosen α but is aware that α' is also perfectly rational, then she's committed to thinking that α will do reliably better than α' . After all, since she knows what α' would think about p given any evidence, we have $\alpha(p|E, \alpha'(p|E) = x) = \alpha(p|E)$. As we saw in the first argument, this steadfast position requires Alice to think her credence will render α' conditionally insensitive to truth-values. But there's no escaping this self-favoring, since otherwise we couldn't have disagreeing yet rational credence functions! Any rational agent with knowledge of another alternative rational credence function has to think that her credence function is superior.

So, Alice thinks her choice of prior was better than a rational prior selected at random. If permissivism is true, this belief is rational. The

²⁰ Again, see (Cohen 2013) for an exception.

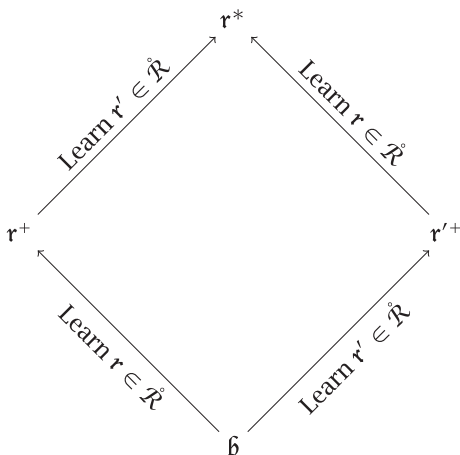


Fig. 3. Collapse

permissivist *qua* theorist should in turn think that rational agents tend to do better than random. Otherwise, she's in an awkward situation: on the one hand, she advises agents to expect that their selection of rational prior would beat one selected at random, but on the other, she herself doesn't really think that's true.

I won't here give a full-throated defense of the view that rational *agents* can reliably do better than random rationality, but I do think there are some reasons not to dismiss this view out of hand.

4.1. Beating Unique Rationality

First, let's see how agents might beat unique rationality. We above argued that if UNIQUE is true, then something like NEW RATREF was correct. In general, if you discover what the uniquely rational credence function thinks about p on your evidence, you should listen to it.

Learning what's rational will tend to render the opinions of other agents conditionally insensitive to whether p . If I know that my credence in p is the rational one and then find out John has the same evidence and disagrees, I should just ignore John. So, we have $\Pr(J = x | \mathfrak{r}^\circ = \mathfrak{r}, p) = \Pr(J = x | \mathfrak{r}^\circ = \mathfrak{r}, \neg p)$.

Usually, that's right. But sometimes agents can reliably do bit better than unique rationality. Consider the following case:

MAX & EVE: Max and Eve are two young but precocious children. Max was born and is aware that he was born with the uniquely most epistemically rational urprior and has always updated by conditionalization. Unfortunately for Max, he still doesn't have an especially accurate credence function, since he hasn't acquired much information about matters like human psychology and the physics of medium-sized dry goods. Eve, on the other hand, doesn't

have the most rational urprior and instead is unjustifiably opinionated about these matters. Her creator, Evolution, knew lots of stuff about the actual world and deviously programmed her—unbeknownst to either child—to have an accurate but unjustified urprior in certain important physical and psychological facts. Max and Eve meet every day on the playground to bet their lunch money on things like how other children and physical objects will behave. To make matters fair, they share all their background evidence. Max initially scoffs at Eve’s brazenly extremal credences, but over time he notices that she has been winning lots of his money.

In this case, Max eventually realizes that Eve is somehow specially sensitive to the truth-values of certain kinds of propositions. Even though he’s rational, she keeps winning money off of him, since she was pre-programmed with a credence function that was more accurate over the relevant domain than the rational one. Her repeated success eventually tips Max off to this fact. Max doesn’t yet know about evolution, but he can infer that Eve’s credence has some *causal* connection to how the world really is, and he should consider her credence to be genuinely informative. Even though Eve lacks any *reason* for her opinion, and Max knows this, the mere fact that she holds it is evidentially important for Max. So, the rational thing for him to do is to start deferring more and more to what Eve thinks.

Note that no analogous case is possible with chance. Suppose a coin will be flipped 1,000 times, and Max is certain that the coin is fair. That is, $\mathfrak{m}(\text{Ch}(H_i) = .5) = 1$, where \mathfrak{m} is Max’s credence function, Ch is the chance function, and H_i is the proposition that the coin will land heads on the i^{th} flip. If the coin actually lands heads the first 999 times, Max—if he’s really certain of the chances—will still have credence .5 it will land heads the next time. Moreover, even if Eve has predicted that the coin will land heads each time, he still will ignore Eve and just consider her a very lucky guesser. While it’s (physically) possible to be sensitive to truth-values in a way that rationality isn’t, it’s not possible to be sensitive to truth-values in a way that *chance* isn’t.²¹ Rationality doesn’t account for all the ways one’s credence could be tied into truth-values.

MAX & EVE looks *recherché*, but it highlights an empirical fact about how we track truth. We really are given a head-start on some propositions—like the ones in the MAX & EVE case—because of our evolutionary history and the causal origins of our doxastic states. If UNIQUE is true, the rational credence function surely won’t be as opinionated as Eve is about human psychology before it gets lots of evidence. So, even for agents like

²¹ For a more in depth discussion of how chance is maximally sensitive to truth-values, see Joyce (2007).

us, beating rationality reliably—at least in limited domains and for a limited amount of time—is indeed not only possible but routine.

4.2. *Beating Permissive Rationality*

In the cases that appear permissive, we arrive at our personal verdicts based on extra-rational factors like hunches and intuitions since reason alone under-determines our choice. The question is whether these extra rational factors can be thought to add reliability. That is, should we think that these factors somehow lead agents to form more accurate credences than they would if they'd chosen a rational prior at random?

It is often reasonable to think that some people just have a good nose for sussing out more accurate rational views from less accurate—but still rational—views. Some gamblers really do have a knack for picking horses, even though they may not exactly be able to articulate why. Some weather persons really can do better than others when it comes to predicting rain. There may simply be no argument for why the gigabytes of data support a credence .8 over .7 even if there's compelling reason not to have a credence of .1. The reasons have simply run out, but some forecasters just are better at truth-tracking than others.

On this view, Alice can take the fact that she's drawn to credence function α as genuine evidence that bears on how accurate she should expect α to be. She knows not just that α is a rational credence function to have, but that a rational agent was—for some reason or other—drawn to incorporate it in her cognitive architecture. So, that at least gives some *prima facie* reason for her to actually adopt α over the alternatives.

Information about the causal origin of Alice's credence function can also be undermining. If Alice discovers she was programmed to have a randomly selected rational prior, she now has no reason to prefer α to other rational credence functions even though that's the one she started with.

We're therefore able to accept PERM EXPERT and the collapse results that come with it (if we like) while maintaining that even rationally omniscient agents can have different credence functions on the same evidence. The reason: we can maintain that

$$\mathbf{r}(p|\mathbf{r}, \mathbf{r}' \in \hat{\mathcal{R}}) = \mathbf{r}'(p|\mathbf{r}, \mathbf{r}' \in \hat{\mathcal{R}})$$

while also maintaining

$$\mathbf{r}(p|\mathbf{r}, \mathbf{r}' \in \hat{\mathcal{R}}, \text{ Alice chose } \mathbf{r}) \neq \mathbf{r}'(p|\mathbf{r}, \mathbf{r}' \in \hat{\mathcal{R}}, \text{ Alice chose } \mathbf{r}')$$

Thus, rational agents can favor their prior over the alternatives if they think they got them in a better-than-random way. This extra bit of information

concerning which probability functions are realized can break the symmetry that was necessary for the collapse argument to go through.

In cases of disagreement with another flesh-and-blood *rational agent*, there won't be any simple way to resolve the dispute. The right response will depend on what the agents think about themselves as truth-measurement devices. In cases in which Alice doesn't have specific information about how she ended up with her prior or about how Bob ended up with his, she doesn't have much reason to favor herself. If she knows that one of the agents has a better track record than the other, she should take that information into account when deciding how to update. What exactly she should do will depend on the specifics of the situation.

It's still unclear whether two rational agents who share the same evidence can agree to disagree. We won't try to resolve the issue here, as the right answer will likely turn on subtle views about the nature of evidence. How attached Alice should be to her prior should depend in part on information she has about why she has that prior, and the same goes for Bob. So, if they really do share all their background evidence, then some of that evidence is about the origin of their priors, their previous track-record, and so on. Some of this evidence may even be ineffable—how strong a hunch is or why one rational response seemed better than another. Perhaps some room for disagreement is left.²²

Regardless of whether disagreement between rational agents can be maintained, disagreeing agents should generally be conciliatory towards one another. One reason is that they can gain knowledge about what probability functions are rational. More important, however, is that they gain knowledge about the readings of good truth-measurement devices. The right response to disagreement will then depend on the specifics of why they ended up with the prior they did.²³

References

- Ballantyne, N. and E. Coffman (2011). Uniqueness, evidence, and rationality. *Philosophers' Imprint* 11(18), 1–13.
- and — (2012). Conciliationism and uniqueness. *Australasian Journal of Philosophy* 90(4), 657–670.

²² However, even if not, the irrationality of sustained disagreement is not a threat to the permissive view *per se* in the way that the second argument is. Rational agents could still know what other functions are rational without ending up in agreement with them.

²³ Thanks to David Christensen, Branden Fitelson, Preston Greene, Amy Hilty, Amanda MacAskill, Dan Marshall, Richard Pettigrew, Miriam Schoenfeld, Brian Weatherston, Gregory Wheeler, the audience at Lingnan University, and an anonymous referee for helpful comments and discussion. Special thanks to Jason Konek for detailed feedback on multiple drafts. I was supported by the ERC Starting Researcher Grant 'Epistemic Utility Theory' during some of the work on this paper.

- Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press.
- Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review* 116(2), 187–217.
- (2010). Rational reflection. *Philosophical Perspectives* 24, 121–140.
- (2014, Online first). Conciliation, uniqueness and rational toxicity. *Noûs*, 1–20.
- Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly* 49, 113–124.
- Cohen, S. (2013). A defense of the (almost) equal weight view. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays*, Chapter 5, pp. 98–120. Oxford University Press.
- Douven, I. (2009). Uniqueness revisited. *American Philosophical Quarterly* 46, 347–361.
- Elga, A. (2007). Reflection and disagreement. *Noûs* 41(3), 478–502.
- (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies* 164, 127–139.
- Enoch, D. (2010). Not just a truthometer: Taking oneself seriously (but not too seriously) in cases of peer disagreement. *Mind* 119(476), 953–997.
- Feldman, R. (2007). Reasonable religious disagreements. In L. Antony (Ed.), *Philosophers without Gods*, pp. 194–214. Oxford University Press.
- Gutting, G. (1982). *Religious Belief and Religious Skepticism*. University of Notre Dame Press.
- Hall, N. (1994). Correcting the guide to objective chance. *Mind* 103, 505–18.
- Horowitz, S. (2013, Online first). Epistemic akrasia. *Noûs* 00(0), 1–27.
- (2014). Immoderately rational. *Philosophical Studies* 167, 41–56.
- James, W. (1979). *The Will to Believe and Other Essays in Popular Philosophy*. Harvard University Press.
- Jaynes, E. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science* 65, 575–603.
- (2007). Epistemic deference: The case of chance. *Proceedings of the Aristotelian Society* 107, 187–206.
- Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne and T. Gendler (Eds.), *Oxford Studies in Epistemology*, Volume 1. Oxford University Press.
- (2010). Peer disagreement and higher order evidence. In R. Feldman and T. Warfield (Eds.), *Disagreement*, pp. 111–174. Oxford University Press.

- (2014). Evidence can be permissive. In M. Steup, J. Turri, and E. Sosa (Eds.), *Contemporary Debates in Epistemology*, pp. 298–311. Wiley-Blackwell.
- Lasonen-Aarnio, M. (2014). New rational reflection and internalism about rationality. Forthcoming in *Oxford Studies in Epistemology*.
- Lewis, D. (1980). A subjectivist's guide to objective chance. In *Studies in Inductive Logic and Probability*, Volume 2, pp. 83–132. University of California Press. Reprinted in *Philosophical Papers*, Volume II, pp. 83–113.
- Rosen, G. (2001). Nominalism, naturalism, epistemic relativism. *Philosophical Perspectives* 15, 60–91.
- Rosenkrantz, R. (1981). *Foundations and Applications of Inductive Probability*. Ridgeview Press.
- Ross, J. (2006). Acceptance and Practical Reason. Ph.D. thesis, Rutgers University.
- Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs* 48(2), 193–218.
- Solomonoff, R. J. (1964a). A formal theory of inductive inference, part 1. *Information and Control* 7(1), 1–22.
- (1964b). A formal theory of inductive inference, part 2. *Information and Control* 7(2), 224–254.
- Thau, M. (1994). Undermining and admissibility. *Mind* 103, 491–504.
- Titelbaum, M. G. (2014). Rationality's fixed point (or: In defense of right reason). In *Oxford Studies in Epistemology*. Oxford University Press.
- van Fraassen, B. (1984). Belief and the will. *Journal of Philosophy* 81, 235–256.
- Weatherson, B. (ms). Do judgments screen evidence? Unpublished manuscript.
- White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives* 19, 445–459.
- (2013). Evidence cannot be permissive. In M. Steup, J. Turri, and E. Sosa (Eds.), *Contemporary Debates in Epistemology*. Wiley-Blackwell.
- Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford University Press.