The Reliabilist and the Bayesian

Bob Beddor

Abstract

If asked to name the most important developments in contemporary epistemology, two plausible answers would be reliabilism and Bayesianism. Despite their prominence, these two developments are rarely brought into conversation with each other. This paper argues that a synthesis can be mutually beneficial. Reliabilism offers the resources for developing a promising new solution to one of the main problems for Bayesianism: the problem of the priors. At the same time, some of the most pressing problems for reliabilism can be solved using Bayesian tools.

1 Two epistemological revolutions

One of the most important developments in epistemology over the last century was the *reliabilist revolution*.[1] For most of its history, epistemology was staunchly internalist. A long tradition stretching from Descartes to the present day has sought to justify an agent's beliefs from the first-person perspective, using only resources available from the armchair. Reliabilists contend that this approach overlooks a crucial dimension of justification. According to reliabilists, justification is a function of reliability, understood as a tendency to produce true beliefs across some class of relevantly similar situations. This is an externalist dimension of justification—one that cannot be discovered from the armchair.

Reliabilism carries significant epistemological advantages. It provides a powerful strategy for combating skepticism. Even if we cannot rule out skeptical hypotheses, our beliefs about the external world can still be justified as long as they are reliably formed and maintained. A second advantage is its reductive promise: reliabilism purports to explain justification in purely non-epistemic terms. It thus offers a way of situating epistemic justification in a naturalistic worldview.[2]

Another major development in epistemology over the last century was the *Bayesian revolution*. Bayesianism begins with the observation that belief is not all-or-nothing: we have higher degrees of belief (credences) in some propositions than in others. Bayesians proceed to provide a simple but powerful theory of the norms governing rational credence. This theory has been developed with a high degree of mathematical precision, and

[1] I borrow this phrase from Williams 2016. For the purposes of this paper, I will construe reliabilism broadly to include other kindred externalist views, such as virtue epistemology, proper functionalism, and various truth-tracking approaches.

[2] See Goldman 1979: 90, who emphasizes reliabilism's promise of a naturalistic reduction of epistemic concepts.

is supported by an impressive battery of arguments. Bayesian approaches have also been influential beyond epistemology, finding fruitful applications in decision theory, cognitive science, and the social sciences.

Both reliabilism and Bayesianism are undeniably important developments. But they are rarely brought into conversation with one another. Standard reliabilist views tell us what it is for a full belief to be justified. But they usually fall silent about what it is for a credence to be justified, or how reliabilist epistemology relates to Bayesian norms.[3]

By contrast, Bayesianism provides an elegant theory of rational credence, but it has primarily been developed from an internalist perspective. The standard Bayesian approach proceeds in two stages. At the first stage, it specifies rational constraints on an agent's *ur-priors*: their credences before they get any evidence about the world. At the second stage, it provides a norm (Conditionalization) that tells agents how they should update their priors in response to evidence. At both stages, the standard Bayesian perspective focuses on how an agent can best achieve certain goals *by their own lights*. For example, Dutch Book arguments show that if you do not abide by Bayesian strictures, you will be susceptible to a loss in utility no matter what the world is like—something that you could discover from the armchair. In a similar vein, expected accuracy arguments purport to show that you will maximize expected accuracy if you adhere to Bayesian norms, where the expectations are calculated using your own credences. For all that these arguments show, your credences might be highly *unreliable*.

In recent years, externalism has made some inroads into Bayesian territory. There is a burgeoning literature on whether Bayesianism is compatible with externalist views of evidence—views on which one can gain p as evidence without being aware that p is part of one's evidence. But these externalist incursions have had a fairly narrow scope. They are only externalist about the evidential inputs to the Bayesian machinery. Discussions of the rational constraints on the priors, and the rational way for agents to update in response to these external inputs, have retained an internalist orientation.[4] In this regard, these discussions differ sharply from the much more thoroughgoing externalism of classic reliabilist views.

[3]There are some important recent exceptions: Dunn 2016, Tang 2016, and Pettigrew 2021 all offer suggestions for how to extend reliabilism to credences. However, none of these authors discuss how their accounts relate to Bayesianism. In fact, all of their proposals conflict with Bayesianism: they all allow that an agent can flout Bayesian norms while having fully justified credences. By contrast, the main goal in this paper is to explain how we can preserve both Bayesianism and reliabilism in a unified framework.

[4]See Schoenfield 2017; Gallow 2021; Das 2022; Medina 2023; Meehan and Zhang 2025. For the most part, these papers focus on how agents should update in response to externalist evidence if they want to maximize expected accuracy. Since the expectations are given by the agent's credences (which might be highly unreliable), this amounts to an internalist approach to external evidence.

A notable exception to this general internalist orientation is Williamson 2000. Williamson champions Bayesian conditionalization as the right response to external evidence, but without trying to provide any justification for this norm, internalist or otherwise. On this picture, conditionalization is a primitive epistemic norm. One of the goals of this paper is to provide an alternative to both of these approaches. Rather than taking Bayesian requirements as primitive, the aim is to derive them from a more general normative framework. But this more general framework is reliabilist, not internalist.

But if we take a step back, the sharp divergence between reliabilist and Bayesian approaches is rather puzzling. Nothing in reliabilism forces one to focus on full belief, rather than credence. And nothing in the Bayesian formalism requires an internalist approach.

The puzzle deepens when when we reflect on the axiological underpinnings of reliabilism and Bayesianism. Both reliabilists and Bayesians often invoke:

Veritism The ultimate source of epistemic value is truth/accuracy. Requirements of epistemic justification and rationality are to be explained in terms of this good.[5]

This paper makes the case for synthesizing reliabilism and Bayesianism. I argue that a synthesis overcomes some of the main problems that arise for each view in isolation. A major problem for Bayesianism is the problem of the priors. One reason why this problem has seemed so intractable is that Bayesians have sought to solve it from an internalist perspective: they have tried to provide constraints on an agent's priors that can be justified from the armchair. Reliabilism offers a fresh perspective on this problem. According to this reliabilist solution, priors are justified insofar as they reliably track features of the external world. And that's not all: I show that we can use reliabilist resources to provide a novel justification for the two core Bayesian norms: Probabilism and Conditionalization.

Reliabilists also stand to gain from a synthesis. A synthesis allows reliabilists to extend their view to credences, and to do so in a way that is consistent with the most explanatorily powerful theory of rational credence currently on offer. A Bayesian twist also helps reliabilists overcome some pressing problems for their view—specifically, problems involving defeat and bootstrapping.

2 Bayesianism and the problem of the priors

2.1 Bayesianism

Bayesianism comes in different forms. In its most minimal form, it makes two normative claims. First, a synchronic requirement:

Probabilism At any time, a rational agent's credences obey the probability axioms.

Second, a diachronic requirement:

Conditionalization A rational agent responds to new evidence by conditionalizing their prior credence function c_{old} on the strongest proposition e that they subsequently learn: $c_{new} = c_{old}(- \mid e) = \frac{c_{old}(- \wedge e)}{c_{old}(e)}$ (when defined).

These two norms provide an elegant and powerful theory of rational credence. However, Bayesianism faces some major challenges. Here I will focus on a particularly pressing challenge: the problem of the priors.

[5] See e.g., Goldman 1999; Pettigrew 2016b.

2.2 The problem of the priors

Bayesianism presupposes that agents have an 'ur-prior'—a credence function that models their initial degrees of belief before they have acquired any evidence. The problem of the priors arises when we ask: Are there any rational constraints on the ur-priors beyond probabilistic coherence?

Subjective Bayesians say *no*: anything goes as far as the ur-priors are concerned (Savage 1972; de Finetti 1974; van Fraassen 1989). Subjective Bayesianism faces a well-known problem: it is overly permissive, allowing patently irrational credences to count as justified. Consider irrational Ira, who assigns credence 1 *ab initio* to a host of absurdities—the moon is made of green cheese, all puppies are controlled by goblins, etc. But Ira is probabilistically coherent: he assigns credence 0 to the negation of all of these absurdities, and to anything entailed by their negation. According to subjective Bayesians, Ira's credences are rational. What's worse, for any counterevidence that he receives that is logically consistent with these absurdities, he is rationally required to remain certain of them (given Conditionalization).

Or consider induction. Suppose Ira has counterinductive priors. The more green emeralds he observes, the *less* confident he will be that the next unobserved emerald is green. Again, subjective Bayesianism declares Ira perfectly rational. This seems wrong.

This problem motivates objective Bayesianism, which insists that there are further constraints on the priors. But what are they?

Some propose a Regularity Requirement, which requires that only necessary truths get credence 1.[6] Others propose the Principal Principle, which requires (roughly) that one's credence in p correspond to one's estimate of the objective chance of p.[7] These constraints are too weak to make much headway on the permissiveness problem for subjective Bayesianism. While Regularity would preclude Ira from being certain that the moon is made out of green cheese, it says nothing about a slightly less opinionated Ira, who has a credence of .99 in this proposition. But, intuitively, such an agent is still irrational.

Another popular constraint is the Principle of Indifference (POI), which requires agents to spread their credences evenly over all hypotheses compatible with their evidence.[8] But the POI faces important problems. Perhaps the most well-known is the problem of multiple partitions, AKA 'Bertrand's paradox': it yields inconsistent results relative to different ways of partitioning the hypothesis space. Even if this technical problem can be overcome, a further worry lurks: skepticism. This problem gets less airtime than the problem of multiple partitions, but—to my mind—it is more fundamental. Before acquiring any empirical evidence about the world, you have no evidence favoring the non-skeptical hypothesis (reality is roughly the way it appears to be) over any number of skeptical hypotheses (you are the victim of an evil demon, or a brain in a vat, or a sim, or a Boltzmann brain). You likewise have have no evidence suggesting that you inhabit a induction-hospitable world,

[6] See e.g., Shimony et al. 1968; Carnap 1971.

[7] Lewis 1980, 1994; Hall 1994.

[8] This principle traces back to Bernoulli 1713 and LaPlace 1814. For defenses, see Keynes 1921, White 2009; Pettigrew 2016a; Williamson 2018; Eva 2019; Liu 2025, among others.

rather than a counterinductive world. So the POI implies you should divide your credences evenly over these hypotheses. But this means even after observing many green emeralds, you should be no more confident that the next unobserved emerald will be green than that it will be some other color.[9]

Consequently, objective Bayesians struggle to specify constraints on the priors that are neither too weak (Regularity, the Principal Principle) nor too strong (the POI). A further problem is what we might call the 'Explanatory Problem'. We don't want to simply give a laundry list of constraints on the priors. We want to provide some story about what justifies these constraints. Why these constraints, and not others? This problem is particularly acute for veritists. Can the objective Bayesian's further norms—whatever they are—be grounded in considerations of truth/accuracy?[10]

Of course, objective Bayesians have much to say about these challenges. I will not go into a detailed discussion of the moves available to objective Bayesians. Rather, I want to offer a new perspective on the problem of the priors by drawing a parallel with a debate in traditional epistemology. This parallel will point us in the direction of a new solution.

2.3 The parallel with external world skepticism

The problem of the priors parallels the problem of external world skepticism. The skeptic asks: 'What justifies your belief that you are sitting in front of a computer, reading this paper?' In response, it is tempting to appeal to your sensory evidence—say, your current experience of sitting at a computer, reading this paper. However, the skeptic will point out that this sensory evidence only speaks in favor of your belief given certain background assumptions—say, that you are awake, that your senses are reliable. What, the skeptic asks, justifies these background assumptions? This challenge is essentially the problem of the priors, where the 'priors' are your background assumptions about the world.[11]

One answer comes from coherentist, who says that your background beliefs are justified provided they 'cohere' together. Coherentism is subjective Bayesianism extended to full belief. And it faces the same problem: it is too permissive. Irrational Ira holds a host of absurd beliefs that cohere together (the moon is made out of green cheese, the moon is made out of dairy products, etc.), without those beliefs being rational or justified.

Another response to external world skepticism, tracing back to Descartes, is to try to justify our background assumptions through some *a priori* argument. The main problem for this project—and for that of its contemporary descendant, objective Bayesianism—is that it is far from clear whether any such *a priori* argument will succeed.

Enter the reliabilist. According to the reliabilist, your background beliefs are justified as long as they are reliably held—that is, formed in a way that tends to track the truth across a sufficiently high proportion of counterfactual circumstances. (See §3.1 for more

[9] See Smithson 2017 for related arguments that the POI leads to skepticism.

[10] Though see Pettigrew 2016b for a wide-ranging attempt to ground constraints on the priors in terms of the goal of accuracy.

[11] The connection between external world skepticism and the problem of the priors has not received much attention, though see Berry 2019 for an exception.

detailed discussion.) Reliabilism provides a simple response to external world skepticism, obviating the need for Cartesian heroics. At the same time, it avoids the permissiveness problem for coherentism. A belief that the moon is made out of green cheese does not reliably track the truth, so does not count as justified.

Partly for this reason, many have been attracted to the reliabilist approach. Surprisingly, however, this approach has never been extended to the problem of the priors. The next section takes on this project.

3 Reliabilist priors

This section explores how to develop a reliabilist constraint on the priors. In order to illustrate the basic approach, I start by explaining reliabilism in a bit more detail (§3.1). I then show how to extend it to provide a rigorous constraint on the priors (§§3.2-3.3), and highlight its advantages over more familiar forms of objective Bayesianism (§§3.4).

3.1 Reliabilism

The central reliabilist idea is that:

Central Reliabilist Idea A belief is justified iff it tends to be true across some domain of relevantly similar situations.

This central idea can be fleshed out in different ways. A first choice point concerns whether to go in for indicator reliabilism or process reliabilism. According to indicator reliabilists, a belief B is justified provided that B (or the ground on which B is based) reliably indicates the truth,[12] According to process reliabilists, a belief B is justified provided that B is the result of a reliable process—i.e., a process that usually produces true beliefs rather than false beliefs.[13] For the moment, we need not decide between these frameworks. The view I will ultimately defend combines elements of both.

Both indicator and process reliabilists invoke a domain of relevantly similar situations. Formally, we can represent this domain using an accessibility relation N: a function from a world w to the set of worlds $N(w)$ that are relevantly similar to w. But how should we interpret this relation?

The most common position holds that the relevantly similar worlds to w are just those which are *nearby* to w. But there are other options. Rather than letting 'N' abbreviate *nearby*, we could let it abbreviate *normal*: we could say that $N(w)$ is the set of worlds where conditions are normal for the exercise of the agent's belief-forming faculties at w.[14]

[12] Versions of indicator reliabilism are defended by Armstrong 1973; Alston 1988; Tang 2016.

[13] See Goldman 1979, 1986; Lyons 2009, among others.

[14] For reliabilist—or more broadly externalist—views that explain the relevant domain of circumstances in terms of normality, see Goldman 1986; Leplin 2012; Graham 2012, 2017, forthcoming; Goodman and Salow 2018; Beddor and Pavese 2020. Different philosophers will offer different ways of unpacking the notion of 'normal conditions'. One option is to explain normality in teleological terms—for example, as the circum-

Even without settling these questions of implementation, we can see how reliabilism offers a solution to skeptical woes. Meet Gwen. Gwen has various foundational beliefs: the external world exists; reality is roughly the way it appears to be, etc. Suppose Gwen is in the good case. At her world (and all nearby worlds), her foundational beliefs are true. Similarly, at all (or almost all) worlds where conditions are normal for the exercise of her belief-forming faculties, these beliefs are also true. Consequently, reliabilists will deem her foundational beliefs justified.

Of course, this reliabilist solution is controversial. One long-standing worry concerns how to evaluate the epistemic credentials of agents in the bad case; this is the 'New Evil Demon Problem' (Cohen 1984). Contrast Gwen with Ben. Ben has the same foundational beliefs as Gwen. But his world is ruled by an evil demon, who has decreed that appearances are systematically misleading. At Ben's world (and all nearby worlds), most of Ben's foundational beliefs are false. But does it follow that Ben's beliefs are unjustified? Some find this a bitter pill to swallow.

By now, reliabilists—and externalists more generally—have ventured a number of responses to this challenge. Perhaps even though Ben's beliefs are not justified, they are *excusable* or *blameless*, since he has no way of discerning his demonic deception from the inside.[15] Or perhaps the right lesson is that we should define N in terms of normal circumstances, rather than nearby worlds (Graham forthcoming). Ben's foundational beliefs are reliable in normal conditions for the exercise of his belief-forming faculties, since these conditions will be demon-free.

A detailed examination of these defensive maneuvers is a task for another paper. For now, I propose that we should accept a promissory note, and take seriously the central reliabilist idea. This central idea is parsimonious, explanatorily powerful, and it offers an elegant approach to combating skepticism, even if there remain tricky questions about how best to handle the New Evil Demon Problem. I turn now to consider how we might extend the central reliabilist idea to the priors.

3.2 Accuracy scores

Reliabilism posits a close connection between justification and truth. But it is at best strained to evaluate credences as true or false. Suppose A has a .9 credence that it is raining (r), whereas B has a .2 credence in r. And suppose it is raining. It seems odd to describe A's credence as true or B's credence as false.

However, credences can be appraised for accuracy. In our example, A's credence in r is more accurate than B's. A natural way to conceptualize the accuracy of a credence is in terms of its *distance from the truth*. The closer your credence is to the truth-value of p, the more accurate your credence in p.

stances where the agent's belief-forming faculties are conducive to fulfilling their functions. See Burge 2003, as well as Graham 2017, forthcoming, who calls this sort of view, 'normal circumstances reliabilism'.

[15] See e.g., Littlejohn forthcoming; Williamson forthcoming; for relevant discussion, see Kelp and Simion 2017; Greco 2021; Ballarini 2022.

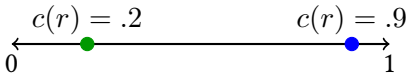

Figure 1: Accuracy as distance from the truth

There are different ways to measure the distance between a credence and the truth. These different measures can be represented using a scoring rule $\mathcal{A}$—a function from a credence function c, a world w, and a set of propositions to a number representing c's degree of accuracy in these propositions at w. For example, one popular scoring rule is the Brier score (Brier 1950). The Brier score takes the inaccuracy of a credence c, in a proposition p, to be the squared Euclidean distance between between $c(p)$ and the truth-value of p. Converting this into a measure of accuracy:

$$\mathcal{B}(c, w, p) = 1 - (T_w(p) - c(p))^2 \tag{1}$$

$$\text{where } T_w(p) = \begin{cases} 1 \text{ if } p \text{ is true at } w \\ 0 \text{ otherwise} \end{cases}$$

To illustrate with our example above: if A assigns a .9 credence to r and r is true, A's credence will get a Brier score of $1 - (1 - .9)^2 = .99$, reflecting a high degree of accuracy. By contrast, B's credence in r will have a Brier score of $1 - (1 - .2)^2 = .36$, reflecting a much lower degree of accuracy.

The Brier score is an instance of a more general class of strictly proper scoring rules:

Strictly Proper Scoring Rules An accuracy score $\mathcal{A}$ is strictly proper iff, for any probabilistic credence function c, the expected accuracy of c (by the lights of c and $\mathcal{A}$), is strictly greater than the expected accuracy of any alternative credence function c' (by the lights of c and $\mathcal{A}$).[16]

We do not need to commit to any particular accuracy score, though I will follow much of the accuracy-based literature in restricting my attention to strictly proper scoring rules. My goal will be to show we can put such a scoring rule to work in developing a reliabilist constraint on the ur-priors.

3.3 From accuracy to reliability

I propose to take seriously the following analogy: *Accuracy stands in the same relation to credence that truth stands in to belief.*

[16] That is, $EA^c(c) \leq EA^c(c')$, with inequality if $c' \neq c$, where:

$$EA^c(c') = \sum_{w \in W} c(w)\mathcal{A}(c', w) \tag{2}$$

For reliabilists, a justified belief is not the same as a true belief. But there is an intimate connection between justification and truth: a justified belief is one that *tends* to be true across some domain of relevantly similar situations. Likewise, a reliablist theory of justified ur-priors should hold that a justified ur-prior is one that tends to be accurate across some domain of relevantly similar situations.

Here is one way of fleshing this out. For simplicity, assume that we have a finite domain of worlds D. Then we can use our preferred accuracy score to define a reliability score over this domain:

Reliability Scores The reliability score $\mathcal{R}$ for some credence c in p, in some domain of situations D, is c's average degree of accuracy with respect to p across D.

Formally:

$$\mathcal{R}(c, p, D) = \frac{1}{|D|} \sum_{w \in \mathcal{D}} \mathcal{A}(c, w, p) \tag{3}$$

To illustrate, take our previous example involving A and B's credences that it is raining. Suppose our domain contains just two worlds, **rainy** and **sunny**, whose names describe their meteorological profiles. And suppose we continue to measure the accuracy of a credence at a world using the Brier score. Then the reliability scores of our agents' credences are given in Table 1.

$c(r)$	$\mathcal{B}(c, \textbf{rainy}, r)$	$\mathcal{B}(c, \textbf{sunny}, r)$	$\mathcal{R}(c, r, \{\textbf{rainy, sunny}\})$
.9	$1-(1-.9)^2 = .99$	$1-(0-.9)^2 = .19$	$(.99+.19)/2 = .59$
.2	$1-(1-.2)^2 = .36$	$1-(0-.2)^2 = .96$	$(.36+.96)/2 = .66$

Table 1: Computing Reliability Scores

Given a measure of reliability, we can impose a reliabilist constraint on the priors:

Reliable Ur-Priors (Local) An ur-prior of c in p is justified at a world w only if $\mathcal{R}(c, p, N(w)) > t$, where t is some threshold.

Recall that $N(w)$ is the set of relevantly similar situations to w. So Reliable Ur-Priors amounts to the claim that your ur-prior in p is justified only if this ur-prior has a sufficently high average accuracy score across the situations that are relevantly similar to yours.

What determines the threshold? This the counterpart of a question that faces all reliabilists: what counts as 'sufficiently reliable'? One option is to follow Goldman 1979 and shrug, pleading vagueness: the concept of justification is vague, so we should not expect a precise answer. However, I think we can say more on this front. Here is a conjecture I find plausible, and which will inform some of what I say going forward: the threshold t is the highest reliability score that the agent is able to achieve, on the topic in question.[17]

[17] Cf. Lasonen-Aarnio forthcoming on 'feasible' dispositions.

This view predicts that different propositions will be indexed to different thresholds. If the agent is capable of achieving a very high reliability score on the topic in question, the threshold will be correspondingly high. If the agent is not capable of achieving a high reliability score on the topic in question—perhaps due to the distribution of worlds in D, or due to the agent's cognitive makeup—then we will set a lower bar.

So far our measure of reliability has been *local*: it evaluates the reliability of an agent's credence in some particular proposition p. But we can generalize this measure to evaluate the reliability of an overall credence function. We can assign accuracy scores not just to credences in particular propositions, but to credence functions. The standard approach is additive: the accuracy of a credence function c is just the sum of the accuracy of the individual credences c assigns to various propositions. That is, where c is defined over an agenda of propositions $\mathcal{F}$:

$$\mathcal{A}(c, w) = \sum_{p \in \mathcal{F}} \mathcal{A}(c, p, w) \tag{4}$$

We can then take the global reliability score of a credence function c to be c's average accuracy score across D (equivalently: the sum of the reliability scores of the individual credences over which c is defined):

Global Reliability Scores The global reliability score $\mathcal{R}$ for some credence function c is c's average degree of accuracy across D.

That is:

$$\mathcal{R}(c, D) = \frac{1}{|D|} \sum_{w \in \mathcal{D}} \mathcal{A}(c, w) = \frac{1}{|D|} \sum_{w \in \mathcal{D}} \sum_{p \in \mathcal{F}} \mathcal{A}(c, w, p) \tag{5}$$

Similarly, we can extend Reliable Ur-Priors (Local) to provide a global constraint:

Reliable Ur-Priors (Global) An agent's ur-prior c is globally justified at w only if c has a sufficiently high global reliability score across the relevantly similar situations (i.e., $\mathcal{R}(c, N(w)) > t$).

I have laid out a simple reliabilist framework. There are various ways of enriching this simple model. So as not to get bogged down in choice points, I will relegate these complications to a footnote.[18] I turn now to the main philosophical payoff of our reliabilist constraint: it solves the problem of the priors.

[18] The reliabilist framework presented in the main text assumes a sharp division between the worlds that are relevantly close to w and those that are not; the latter make no difference to the reliability of one's credences at w. But this assumption could be relaxed. Instead of relying on a fixed domain of relevantly similar worlds, we could adopt a gradational model. On this view, the reliability of a credence function is its weighted accuracy across *all* possible worlds, where the weights are supplied by whatever the externalist cares about (e.g., modal closeness, normality). Formally, for any world w, assume we have a function g_w: $W \to \mathbb{R}$ that assigns numeric weights to worlds based on how well they score, relative to w, along our preferred externalist dimension. (As an example, suppose our externalist dimension is just modal closeness. And suppose we make the 'limit assumption' that for any world w, there will always be a sphere of closest-to-w worlds. Then we could assign

3.4 Solving the problem of the priors

We saw in §2 that subjective Bayesianism is too permissive. Recall irrational Ira, who has credence 1 that the moon is made out of green cheese (m). As long as he is probabilistically coherent, subjective Bayesianism deems his credences rational, which seems wrong.

Reliable Ur-Prior avoids this result. Suppose Ira inhabits a world much like ours. Then m will be false at his world, and at all nearby worlds. Similarly, m will be false at most worlds where conditions are normal for the exercise of his belief-forming faculties. So for any plausible way of specifying the domain of relevantly similar situations, Ira's ur-prior in m will get a very low average accuracy score across the relevantly similar situations, and so will not be justified.

For much the same reason, Reliable Ur-Priors solves the problem of induction. Suppose Ira has counterinductive ur-priors. But let us also continue to suppose that Ira inhabits a world much like ours, where conditions are induction-hospitable. His counterinductive priors will get a low average accuracy score across all relevantly similar worlds, and so will not be justified.

This solution extends smoothly to the new riddle of induction (Goodman 1955). If Ira inhabits a world like ours, then most unobserved emeralds are green, not grue. So if Ira assigns a high prior to the hypothesis that the next unobserved emerald is grue, his gruesome prior will receive a low reliability score.

What if Ira inhabits an abnormal world? Suppose that a benevolent demon manipulates Ira's environment to agree with his seemingly irrational priors. This demon ensures that at Ira's world, the moon is made out of cheese, and that every time Ira observes a green emerald, the proportion of unobserved green emeralds diminishes. Suppose, moreover, that this demon's benevolence is not a mere caprice: at all nearby worlds, the demon also guarantees the accuracy of Ira's priors. Does that render Ira's priors justified?

While this is an interesting question, it is really a question that arises for reliabilists (indeed, externalists) much more generally: How should we evaluate the epistemic status of an intuitively irrational belief/credence formed in abnormally cooperative circumstances?[19] The answer will depend on the details of one's reliabilist theory, and will be closely connected to one's answer to the New Evil Demon Problem. One option is to in-

weights to worlds based on which 'ring' they occupy in a nested similarity sphere centered on w (Smith 2016), with the worlds in the innermost sphere getting assigned the highest weight, and the worlds in the outermost sphere getting the lowest weight.) Assume also that g_w has been normalized, and so:

$$\sum_{v \in W} g_w(v) = 1 \tag{6}$$

We could then define the global (gradational) reliability score for some credence function c at w as c's weighted accuracy score across all possible worlds, where the weights are supplied by g_w:

$$\mathcal{R}(c, w) = \sum_{v \in W} g_w(v)\mathcal{A}(c, v) \tag{7}$$

Our reliabililist constraint on the priors could similarly be revised to require that an ur-prior c is justified at w only if $\mathcal{R}(c, w) > t$.

[19] See Goldman 1979: 100-101 for an early discussion of the benevolent demon objection to reliabilism.

sist that if Ira is fortunate enough to encounter such demonic aid, his beliefs/priors are justified. To the extent that we are inclined to judge otherwise it is because our epistemic judgments are shaped by our assumptions about what sort of beliefs/credences would be accurate in environments like ours. A different option is to appeal to normal circumstances reliabilism: even though Ira's beliefs/credences are reliable at nearby worlds, they are not reliable in normal conditions, since normal conditions are free from demons (evil or benevolent).[20] For our purposes, we can remain noncommittal between these responses.

How does Reliable Ur-Priors compare to objective Bayesianism? Reliable Ur-Priors can be viewed as a new version of objective Bayesianism. But it is a thoroughly externalist version, very different from the main flavors currently on offer. This externalist dimension enables it to overcome the main hurdles facing the more familiar flavors.

We saw that some versions of objective Bayesianism (e.g., those that rely on just Regularity and/or the Principal Principle) are too weak: they struggle to explain what is wrong with a version of irrational Ira who assigns credence .99 to m. Reliable Ur-Priors has no trouble here. Such an agent still gets a low reliability score across any domain of worlds similar to ours.

Turn next to the Principle of Indifference. The first problem for the POI was the problem of multiple partitions. As stated, Reliable Ur-Priors avoids this problem, since it does not make it does not make justification depend on an arbitrary partitioning of the hypothesis space. But we can also put things in a more constructive and irenic vein. We will see in the next section that, for any finite domain D, the ur-prior with the highest reliability score will always be the uniform distribution over D. So Reliable Ur-Priors allows us to recover a restricted form of the Principle of Indifference: the 'best' (most reliable) prior will be evenly spread most fine-grained partition of relevantly similar worlds. Importantly, this is not merely stipulated as the right partition; it emerges as a natural consequence of the reliabilist machinery.[21]

The second (and, in my eyes, more fundamental) problem for the POI is that it leads to skepticism. Reliable Ur-Priors avoids this pitfall. Reliable Ur-Priors does not mandate that rational agents' priors are neutral between the skeptical and non-skeptical hypotheses.Recall Gwen, who inhabits a world where reality is roughly as it appears. Since the nearby worlds—and the normal worlds—will be deception-free, Gwen will get a higher reliability score if she assigns a low prior to the skeptical hypothesis. In this respect, Reliable Ur-Priors is an extension of the more general reliabilist anti-skeptical strategy (§3.1).

Finally, Reliable Ur-Priors avoids the Explanatory Problem for objective Bayesianism. It does not give a long list of arbitrary constraints. It offers just a single constraint. This constraint is grounded in the goal of accuracy, in much the same way that standard reliabilism is grounded in the goal of truth. Consequently, it should be attractive to veritists.

[20] See Graham forthcoming for defense of this response to the benevolent demon problem.

[21] Here it is helpful to compare my framework with the response to White's 2009 response to the multiple partitions problem. White's response is to suggest that there will always be one 'right' way of partitioning the hypothesis space. (See also Liu 2025.) But White does not tell us what the right way is. Reliable Ur-Priors offers an answer.

4 Consequences of the reliabilist constraint

Having developed a reliabilist constraint on the priors, I turn to consider some its implications. Which priors maximize reliability (§4.1)? And how does this constraint relate to Probabilism (§4.2)?

4.1 In search of the most reliable ur-priors

On the view defended here, an ur-prior is justified only if it is sufficiently reliable. This raises a natural question: Which ur-priors are the most reliable?

The answer will depend on what our domain of worlds is like. This is a key feature of the reliabilist approach: the justificatory status of the priors depends on what is going on across relevantly similar worlds. Still, we can make an observation about which priors maximize reliability given some domain:

Fact 1. The Uniform Prior Maximizes Reliability. For any strictly proper scoring rule, the most reliable credence function, relative to some finite domain of worlds D, will be the uniform probability distribution over D (u_D), which assigns credence 0 to every world outside of D, and equal credence to every world in D:

$$u_D(w) = \begin{cases} 1/|D| & \text{if } w \in D \\ 0 & \text{otherwise} \end{cases}$$

Before walking through the proof, it may be helpful to give an intuitive sense for why this holds. To do so, let us consider different ways the truth-value of p might be distributed across D. For any such distribution, we can plot the reliability scores of different credences one could take towards p.

First, suppose p's truth-value is invariant across D: p is true in all worlds in D or none of them. Either way, the reliability score for a given credence in p will just be its accuracy score at any world in D. If p is true at every world in D, then the credence in p with the worst reliability score is credence 0, and the credence in p with the best reliability score is credence 1. This is plotted in the blue curve in Fig. 2, where the reliability score is calculated using the average Brier score. If p is true in none of the worlds in D, then the situation is reversed: the credence in p with the lowest reliability score is credence 1, and the credence with the highest reliability score in 0, as reflected in the red curve in Fig. 2. Either way, the credence with the highest reliability score equals the fraction of worlds in D where p is true (1 or 0).

What if p's truth-value varies across D? Then no credence in p will have a perfect reliability score. But the credence in p with the best achievable reliability score will always equal the fraction of worlds in D where p is true. For example, the green curve in Fig. 3 plots the reliability scores of the various credences an agent could adopt towards p when p is true in half of the worlds in D (where, again, reliability is calculated as the average Brier score). In this scenario, the credence that achieves the best possible reliability score

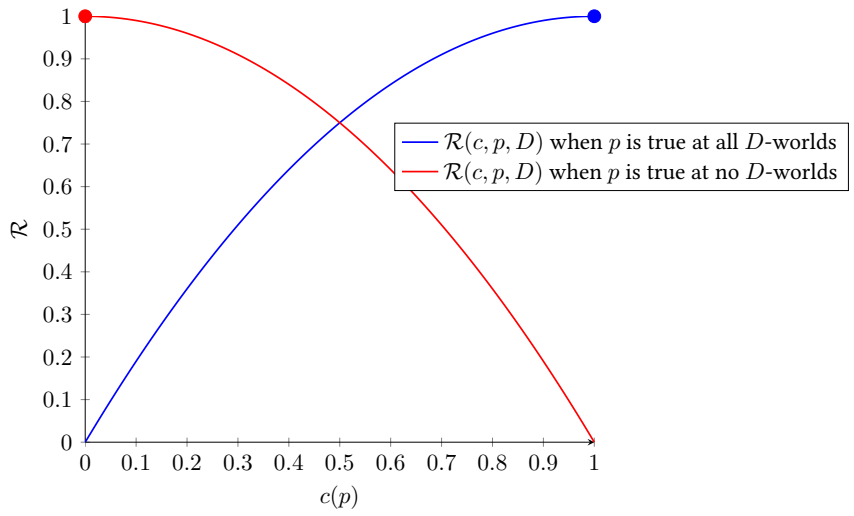

Figure 2: Reliability scores when p's truth-value is invariant across D

is .5. Similarly, the blue curve in Fig. 3 plots the reliability scores for different credences in p when p is true in 75% of the worlds in D. The credence that achieves the best possible reliability score in this scenario is .75.

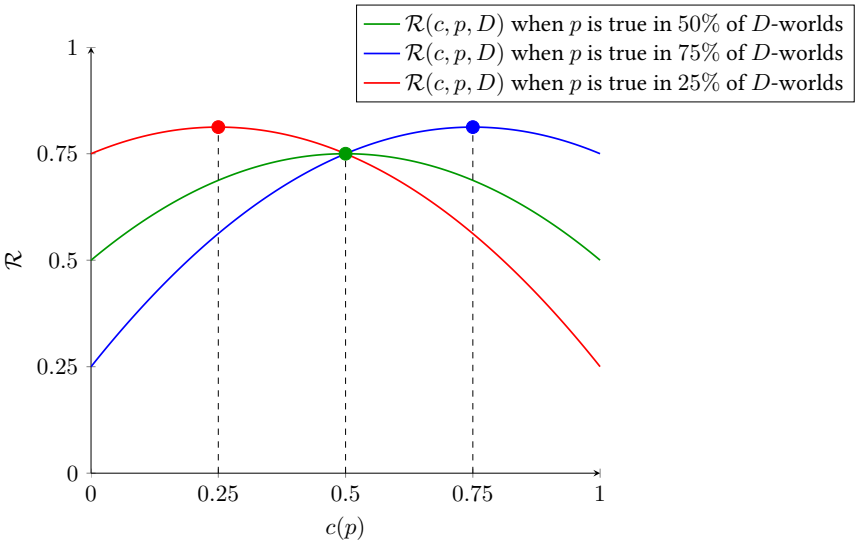

Figure 3: Reliability scores when p's truth-value varies across D

To see the full range of possibilities simultaneously, we can use a three dimensional plot (Fig. 4), where the x axis represents the different credences an agent can have towards p, the y axis represents the different fractions of worlds in D were p is true, and

the z axis represents reliability scores. Note that for any particular fraction n of p-worlds, the credence in p with the highest possible reliability score is also n. The only credence function that has this property is the uniform distribution over D.

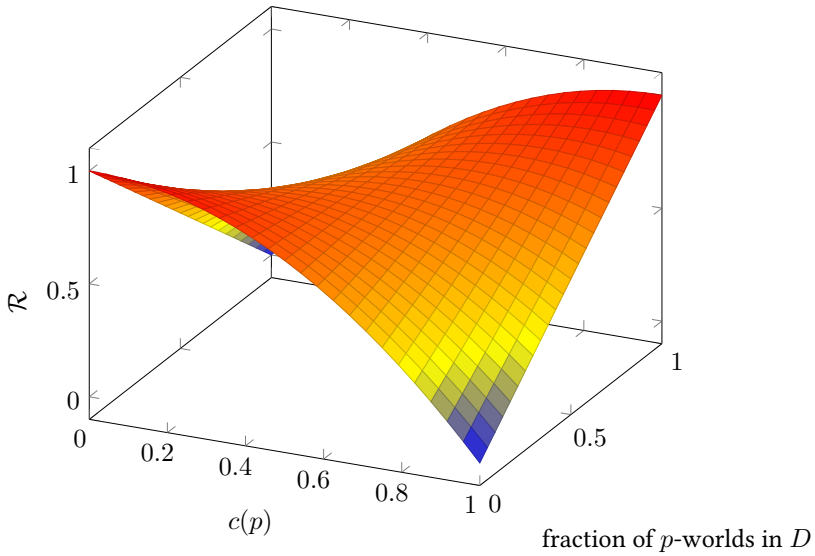

Figure 4: Reliability scores as a function of the fraction of p-worlds in D

Having illustrated the basic idea, I will now offer a general proof of why Fact 1 holds for all strictly proper scoring rules, not just the Brier score.

Proof of Fact 1. The expected accuracy of a credence function c, as evaluated by u_D, is calculated in the usual way:

$$EA^{u_D}(c) = \sum_{w \in W} u_D(w)\mathcal{A}(c, w) \tag{8}$$

Since u_D assigns credence 0 to any world not in D, and credence $1/|D|$ to every world in D, we get:

$$EA^{u_D}(c) = \sum_{w \in D} u_D(w)\mathcal{A}(c, w) = \frac{1}{|D|} \sum_{w \in D} \mathcal{A}(c, w) \tag{9}$$

And this is equivalent to c's global reliability score with respect to D (Global Reliability Scores). So:

Reliability-Expected Accuracy Bridge For any finite domain D, the reliability score of a credence function c, relative to D, is c's expected accuracy from the perspective of u_D:

$$\mathcal{R}(c, D) = EA^{u_D}(c) \tag{10}$$

The next step in the proof is to appeal to the signature property of strictly proper scoring rules. For any strictly proper scoring rule, the expected accuracy of a probabilistic credence function c, when calculated using c, is strictly greater than the expected accuracy of any alternative credence function: $EA^{c'}(c) \leq EA^{c'}(c')$, with inequality if $c' \neq c$. As a special case where expected accuracy is calculated using u_D, it follows that if our accuracy score is strictly proper, then $EA^{u_D}(c) \leq EA^{u_D}(u_D)$, with inequality if $u_D \neq c$. So $\mathcal{R}(c, D) \leq \mathcal{R}(u_D, D)$, with inequality if $u_D \neq c$ (by Reliability-Expected Accuracy Bridge). Consequently, the credence function that maximizes the reliability score, relative to any domain of worlds D, will be the uniform distribution over D (u_D).[22] □

Does this mean that if an agent's priors do not agree with the uniform distribution over the relevantly similar worlds, their priors are unjustified? Not necessarily. Reliable Ur-Priors only requires that the agent's credences are *sufficiently* reliable. I proposed that the bar for sufficient reliability is constrained by ability: it is highest reliability score the agent is able to attain. Suppose someone is not capable of conforming their credences to the uniform distribution—say, because of limited powers of logical discernment. Then their credences might still be justified. However, conforming one's credences to the uniform distribution over the relevant domain remains the reliabilist ideal. While ordinary agents might fall shy of this ideal while still retaining justification, their ur-priors would be epistemically suboptimal.

4.2 The relation to Probabilism

What is the relation between Reliable Ur-Priors and Probabilism? We saw that probabilistic coherence is not sufficient for reliability: irrational Ira is probabilistically coherent, but his ur-prior in m gets the lowest reliability score. Does the converse entailment hold? Are all reliable ur-priors probabilistically coherent?

At the local level, *no*. Reliable Ur-Priors (Local) only cares about the accuracy of your credence in some particular proposition across the relevantly similar situations. Coherence is a global property: it depends on how your credences in different propositions hang together. Suppose incoherent Ina assigns credence 1 to p, and credence .1 to $\neg p$. And suppose p is true at every world. Then Ina's credence in p gets the highest possible reliability score, and so satisfies Reliabilist-Ur Priors (Local). But she violates Probabilism.

This is a feature rather than a bug. Probabilism is a very demanding requirement, as revealed by the problem of logical omniscience. Even if we want to insist that Probabilism is a requirement of rationality, we still want to have something positive to say about the many agents who fail to live up to it. Reliable Ur-Priors offers one way to pay epistemic

[22] In fn.18 I noted that we could instead adopt a gradational model of reliability, according to which the reliability of a credence c is c's weighted average accuracy score, where the weight g_w is supplied by some externalist ordering (e.g., modal closeness, normality). On this model, a prior maximizes reliability at some world w if and only if it agrees the externalist weighting (i.e., $\forall v : c(v) = g_w(v)$). This will be the uniform prior over D in the special case where g_w assigns weight 0 to every world outside of D and equal weight to every world in D.

tribute to such agents. Even if an agent is incoherent, we can still positively evaluate their individual credences in particular propositions from a reliabilist perspective.

At the same time, it would be too hasty to conclude that there is no connection between reliability and Probabilism. Take incoherent Ina. While her credence in p achieves the highest possible reliability score, her .1 credence in $\neg p$ does not. If we calculate accuracy using the Brier score, her credence in $\neg p$ gets a reliability score of .99. This is high, but not as high as it would be she assigned credence 0 to $\neg p$.

We can state a more general result. Earlier we observed that reliable ur-prior, relative to any finite domain D, is the uniform probability distribution over D (Fact 1). And the uniform probability distribution is, of course, probabilistically coherent. It follows that:

Fact 2. Incoherence is Never Fully Reliable. If accuracy is measured using a strictly proper scoring rule, then for any finite domain of worlds D, the credence function with the highest achievable global reliability score relative to D will always be probabilistic.

So there is a connection between our reliabilist constraint and Probabilism after all. While an incoherent agent's credence in some particular proposition might still attain the highest possible reliability score, their overall credence function cannot. Coherence is necessary for maximal global reliability.[23]

This does not necessarily mean that an incoherent agent's ur-priors are unjustified. Incoherent ur-priors might still qualify as *sufficiently* reliable, provided that they attain the highest reliability score that the agent is genuinely capable of achieving. (Imagine that the ideal of logical omniscience is beyond our agent's cognitive reach.) But even then, our agent's incoherent ur-priors will be epistemically suboptimal, and not just from an internalist perspective. They are also suboptimal from a reliabilist standpoint.

Some might think this is too quick. My argument that incoherence is never fully reliable assumed that our accuracy score is strictly proper. This is a standard assumption in the accuracy literature. However, our objector may observe, the standard rationale for focusing on strictly proper scoring rules is broadly internalist. It goes like this. Suppose that, by the lights of your current credence function c, some other credence function c' had higher expected accuracy than yours. Then your credences would be in a certain sense self-undermining: if your goal was to maximize expected accuracy, you would do better to abandon c in favor of c'.[24] This rationale focuses on how the agent can *expect* to do best, *vis-à-vis* the goal of achieving accuracy. In this respect, it is an internalist argument. What place does it have in a reliabilist framework?

[23] We could also reach the same result through a different path. A famous result in the accuracy-first literature shows that, for any strictly proper scoring rule, any incoherent credence function c is strongly accuracy-dominated by some probabilistically coherent credence function c', in the sense that at every world c' strictly more accurate than c. By contrast, no probabilistic coherent credence is even weakly accuracy-dominated by an incoherent credence function. (See de Finetti 1974; Joyce 1998; Predd et al. 2009; Pettigrew 2016b.) This gives us another route to the conclusion that incoherence is never fully reliable. After all, suppose that c is not probabilistically coherent. Then there is some probabilistic credence function c' that is more accurate than c at every world (by the accuracy dominance result). So for any domain of worlds D, c' will have a higher average accuracy score across D than c does. So $\mathcal{R}(c, D) < \mathcal{R}(c', D)$ (by Global Reliability Scores).

[24] See e.g., Oddie 1997; Greaves and Wallace 2006; Horowitz 2013.

However, a synthesis of reliabilism and Bayesianism need not dispense with internalist arguments altogether. Rather, it might embrace a combination of internalist and reliabilist considerations. We could agree that agents should strive to maximize expected accuracy, by their own lights. And so we can avail ourselves of the standard internalist rationale for strictly proper scoring rules. But we might also insist epistemic justification also has an externalist dimension. We should not just care about maximizing *expected* accuracy. We should also care about whether one's credences tend, in fact, to be accurate. Here is where the reliabilist element kicks in.

5 Reliabilist posteriors

5.1 Towards a two-stage model

I have advanced a reliabilist constraint on the priors. However, the normative constraints on the priors are just part of the Bayesian picture. Rational agents also update their credences in response to evidence. How should reliabilists evaluate posterior credences?

An initially tempting thought is that we could just apply our reliabilist constraint on the priors to the posteriors. On this view, an agent's posterior credence is justified at w only if it receives a sufficiently high reliability score across $N(w)$. But there are two problems with this proposal.

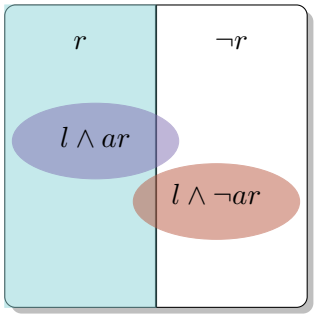

Figure 5: Looking for rain

First, it yields the wrong results in cases where someone starts with perfectly reliable priors, but subsequently updates these priors based on new evidence. Consider an elaboration our earlier model (Fig. 5). As before, half of the relevantly similar worlds are rainy. You have been inside all day, but you could glance out the window. Your visual appearances are generally reliable: at 90% of the relevantly similar where you look outside (l) and it appears to be raining (ar), it is raining. Similarly, at 90% of the the relevantly similar worlds where you look outside and it does not appear to be raining ($\neg ar$), it is not raining. Suppose your priors agree with the uniform probability distribution over $N(w)$, and so you have a .5 credence in r. You now glance out the window, and see that it appears to be raining. Intuitively, you should increase your credence in r. This is not just

an internalist intuition; reliabilists should agree, given the reliability of your vision. But if you increase your credence in r, your posteriors will no longer agree with the uniform distribution over $N(w)$, and hence will become less reliable relative to $N(w)$.[25]

The second problem is related. Our reliabilist constraint on the ur-priors is purely synchronic: it only looks at the accuracy of an agent's credences across some domain. It does not pay attention to how the agent arrived at those credences. When it comes to the ur-priors, this makes sense: the ur-priors are not the result of any learning process. But when it comes to articulating a constraint on the posteriors, this feature becomes a serious limitation. Justification also has an *ex post* or doxastic dimension, which depends on the way in which an agent arrived at their beliefs.[26] Suppose two detectives have the same evidence that the butler committed the crime. Suppose both also believe the butler is guilty. But detective A arrived at their belief through careful consideration of the evidence, whereas detective B arrived at their belief by flipping a coin. There is an important justificatory difference between their beliefs. An adequate epistemology should explain this difference. Traditional process reliabilism has a simple explanation: the first detective used a reliable process; the second did not. But our reliabilist constraint on the ur-priors is not a *process* reliabilist constraint, and so it cannot avail itself of this explanation.

The contrast with process reliabilism suggests a natural path forward. In order to account for justified posteriors, we should take into account the reliability of the updating process that the agent employed when arriving at those posteriors.

5.2 Reliable updates

I will work with a simple model that represents an updating process as a function $\mathcal{P}$ from a world w to a credence function $\mathcal{P}_w$. Suppose your update your pluvial credences based on looking out the window. Your updating process modeled as a function that takes as input some world w, and produces as output a posterior credence that assigns a high credence to r if it appears to be raining at w, and produces a low credence in r otherwise.[27]

How should we score updating processes for reliability? While different approaches are possible, here is a natural option:

Reliability Scores for Updating Processes The reliability score $\mathcal{R}$ for some updating process $\mathcal{P}$, in some domain of situations D, is the average accuracy scores of the credence functions that $\mathcal{P}$ produces across D.

That is:

[25] Perhaps, some might suggest, we should evaluate your posteriors relative to a restricted set of relevantly similar worlds. Rather than evaluating them for reliability relative to $N(w)$, we should evaluate them relative to the worlds in $N(w)$ where you undergo the same experience ($N(w) \cap ar$). While this seems plausible, there is a question whether we can give a principled justification for this restriction on reliabilist grounds. The reliabilist framework I develop below avoids making any such stipulative restrictions.

[26] See e.g., Firth 1978. For an overview, see Silva and Oliveira 2023.

[27] Really there are a bunch of different updating functions here, which differ in terms of how much weight they give to the visual appearances. More on this shortly.

$$\mathcal{R}(\mathcal{P}, D) = \frac{1}{|D|} \sum_{w \in \mathcal{D}} \mathcal{A}(\mathcal{P}_w, w) \tag{11}$$

We could then propose:

Reliable Posteriors An agent's posterior credence is justified only if it results from using an updating process that is sufficiently reliable across the relevantly similar situations.[28]

Here too, we face the question: What counts as 'sufficiently' reliable? Here too, a plausible answer is to appeal to *availability*. The threshold for reliability is the reliability score of the most reliable process available to the agent, in the sense that they are able to use this process in their circumstances. (More on how to unpack availability shortly.)

A view along these lines overcomes the problems noted in §5.1. First, it explains why you are justified in updating your .5 credence in r on the basis of your visual experience (Fig. 5). Since visual appearances are reliable in this domain, the process of adjusting your credence in response to the visual appearances will be more reliable than the 'dogmatic' process of sticking with your prior credence in r regardless of what you see.

On to the second problem: updating processes are abstract characterizations of psychologically real processes going on within the agent. So our reliabilist constraint on the posteriors makes the justificatory status of a posterior credence depend on *how* the agent arrived at that credence.

5.3 Reliable updating and availability

Which updating processes are the most reliable?

The most reliable *logically possible* updating process will always be the *omniscient update* which, for any world w, assigns credence 1 to all and only the true propositions at w. Sadly, however, the omniscient update is not available to creatures like us.

Why not? Because we are constrained by our evidence. One way of fleshing out this explanation, due to Greaves and Wallace 2006, appeals a notion of *evidential constancy*. Let a learning situation $\mathcal{L}$ be a function from a world w to the (strongest) proposition $\mathcal{L}(w)$ that the agent learns in that situation. For example, suppose at world w_1 you look outside and see that it appears to be raining. Then $\mathcal{L}(w_1)$ might be the proposition: *I have looked outside and it appears to be raining*. Say that an updating process $\mathcal{P}$ is evidentially constant iff for any learning situation where the agent learns the very same proposition, $\mathcal{P}$ outputs the same posteriors. Following Greaves and Wallace, we can propose:

Evidential Constancy An updating process $\mathcal{P}$ is only available to an agent in a learning situation if $\mathcal{P}$ is evidentially constant (i.e., whenever $\mathcal{L}(w) = \mathcal{L}(w')$, $\mathcal{P}(w) = \mathcal{P}(w')$).

[28] That is: $\mathcal{R}(\mathcal{P}, N(w)) > t$, where $\mathcal{P}$ is whatever updating process the agent employed.

Some might worry that this constraint smuggles an internalist element into our framework. But this worry is misplaced, for two reasons. First, as noted in §4.2, a synthesis of Bayesianism and reliabilism need not eschew internalism altogether. Rather, it could combine internalist and externalist considerations. Second, even staunch reliabilists rely on a similar constraint. Process reliabilists understand processes as functions from input states of the agent to doxastic outputs (Goldman 1979). Evidential Constancy encodes a similar idea, where the input to the process is the strongest propositioned learned.

Evidential Constancy rules out the omniscient updating process. Can we say anything about which evidentially constant updating processes are the most reliable?

Yes. We can show that when an agent's prior c maximizes reliability relative to some domain D, then in an important class of learning situations the most reliable evidentially constant updating process will be to conditionalize c on the agent's evidence. This gives us a new, distinctly reliabilist argument for Conditionalization.

5.4 When Conditionalization is the most reliable available process

My argument builds on a famous *internalist* defense of conditionalization, due to Greaves and Wallace 2006. Greaves and Wallace assume (as will I continue to do) that accuracy is measured using a strictly proper scoring rule. They show:

Fact 3. Expectation Theorem (Greaves and Wallace) If a learning situation $\mathcal{L}$ is factive and partitional, then for any probabilistic prior c, the evidentially constant updating process that maximizes expected accuracy (by the lights of c) will conditionalize c on $\mathcal{L}$ (i.e., for every w, $\mathcal{P}_w(-) = c(- \mid \mathcal{L}(w))$.[29]

This is an internalist justification for Conditionalization, and it has been historically championed as such. The idea is that any agent will, by their own lights, expect to do best if they conditionalize.[30] But for all that has been said, the agent's own lights might be systematically unreliable.[31]

However, we can repurpose this internalist result in service of a reliabilist conclusion. Recall our Reliability-Expected Accuracy Bridge, according to which the reliability of score of a credence function c, relative to some domain D, equals the expected accuracy of c, by the lights of the uniform distribution over D (u_D). A similar bridge principle holds for updating processes:

Reliability-Expected Accuracy Bridge for Updates The reliability of an updating process $\mathcal{P}$, relative to some domain D, is the expected accuracy of $\mathcal{P}$ from the perspective of the uniform probability distribution over D (u_D).[32]

[29] For the proof, see Greaves and Wallace 2006.

[30] See Leitgeb and Pettigrew 2010; Easwaran 2013 for related, and similarly internalist, expected accuracy-based arguments for Conditionalization.

[31] See Douven 2013 for a criticism along these lines.

[32] *Proof sketch:* The expected accuracy of an updating process $\mathcal{P}$ is just the expected accuracy of adopting the posterior credences that $\mathcal{P}$ produces:

Say an updating process $\mathcal{P}$ is *u_D-conditionalizing* on $\mathcal{L}$ iff it agrees with the result of conditionalizing u_D on $\mathcal{L}$: $\mathcal{P}(- \mid \mathcal{L}) = u_D(- \mid \mathcal{L})$. As a special case of Fact 3: for any factive and partitional learning situation $\mathcal{L}$, the evidentially constant updating process that maximizes expected accuracy, by the lights of u_D, will be u_D-conditionalizing on $\mathcal{L}$. Given the Reliability-Expected Accuracy Bridge for Updates, we can conclude:

Fact 4. Conditionalization-Reliability Connection. If $\mathcal{L}$ is factive and partitional, the most reliable evidentially constant updating process $\mathcal{P}$, relative to some domain D, will be u_D-conditionalizing on $\mathcal{L}$.

We also know that, for any finite domain D, the most reliable prior to have, relative to D, is u_D (Fact 1). So from Facts 1 and 4 we can conclude:

Fact 5. Conditionalization Preserves Maximal Reliability. For any factive and partitional learning situation $\mathcal{L}$, if an agent's prior c achieves the highest possible reliability score with respect to D, the most reliable evidentially constant updating process will conditionalize c on $\mathcal{L}$.

To illustrate, recall our example depicted in Fig. 5. Your prior is uniformly distributed over the relevantly similar situations. If you glance outside, you will either learn that you have looked outside and it appears to be raining ($l \wedge ar$) or you will learn that you have looked outside and it does not appear to be raining ($l \wedge \neg ar$). Suppose also that your evidence is partitional: if you learn $l \wedge ar$, you also learn that you this, and similarly if you learn $l \wedge \neg ar$. Fact 5 tells us that the most reliable evidentially constant updating process will be to conditionalize your priors on what you learn when you look outside. So you will adopt a .9 credence in r if you look outside and it appears to be raining.[33]

Does this mean that if you do not conditionalize in factive and partitional learning situations, your posteriors are unjustified? Not necessarily. While evidential constancy is a plausible constraint on availability, it may not be the only constraint. As critics of Bayesianism emphasize, conditionalization is computationally difficult and expensive. Given our limited memory and computational abilities, it is doubtful whether we are genuinely able to update by conditionalization in all circumstances. Research in cognitive science suggests that humans use a variety of heuristics in reasoning—heuristics that approximate conditionalization in certain contexts, while diverging from it in others. If some

---

$$EA^c(\mathcal{P}) = \sum_{w \in W} c(w)\mathcal{A}(\mathcal{P}_w, w) \tag{12}$$

As a special case, the expected accuracy of $\mathcal{P}$, as evaluated from the perspective of the uniform distribution over D (u_D), is given by:

$$EA^{u_D}(\mathcal{P}) = \sum_{w \in W} u_D(w)\mathcal{A}(\mathcal{P}_w, w) = \frac{1}{|D|} \sum_{w \in D} \mathcal{A}(\mathcal{P}_w, w) \tag{13}$$

This quantity is simply our earlier definition of the reliability of an updating process. So the reliability of $\mathcal{P}$, relative to D, is the expected accuracy of P by the lights of u_D ($\mathcal{R}(\mathcal{P}, D) = EA^{u_D}(\mathcal{P})$). □

[33] Since at 90% of the worlds in $N(w)$ where you look outside and it appears to be raining, it is in fact raining, and so $u_{N(w)}(r \mid l \wedge ar) = .9$).

such heuristic is the most reliable process that an agent is genuinely capable of using, then it could qualify as sufficiently reliable.

5.5 Bayesianism is (sometimes) the reliabilist ideal

I have argued that conditionalization is not just the optimal updating process from an internalist perspective. Under certain conditions, it is also optimal from a reliabilist perspective. If your priors meets the reliabilist ideal, then in factive and partitional learning situations the most reliable available updating process will conditionalize your priors on your total evidence. So if you update by some alternative process in such situations, then you fell short of the reliabilist ideal in some respect. Either (i) your priors did not have the highest achievable reliability score, or (ii) you were not updating using the most reliable available updating process.

In §4.2 I argued that Probabilism is also the reliabilist ideal: if you are not probabilistically coherent, you are less than fully reliable (Fact 2). By combining this result with my reliabilist argument for conditionalization, we get a new—and distinctly externalist—vindication of the two central Bayesian norms.

Still, we should be careful not to overstate this conclusion. My reliabilist vindication of Conditionalization is restricted to specific conditions, which are worth unpacking…

5.6 Reliably updating on unreliable priors

First, our argument that conditionalization is the most reliable available updating process only applies when the agent's priors receive the highest reliability score. What if their priors are less than perfectly reliable?

Fact 4 establishes that in factive and partitional learning situations, the most reliable evidentially constant updating process will always be u_D-conditionalizing—that is, it will always conditionalize the uniform distribution u_D on the agent's evidence. This point holds regardless of whether the agent's priors agrees with u_D.

Here is one way to appreciate this point. For any domain of relevantly similar situations D, the reliabilist's advice for how to update in factive and partitional learning situations is straightforward: update by conditionalizing a privileged prior on your total evidence. What is the privileged prior? Answer: the most reliable prior, relative to D. From Fact 1, we know that this will always be the uniform prior over D. Consequently, in factive and partitional learning situations, the most reliable evidentially constant updating process will always be a conditionalizing process—that is, it will always agree with the result of conditionalizing a privileged prior on your total evidence. But the privileged prior will only be *your* prior if your prior agrees with u_D.

Now, as long as your priors are close to maximally reliable—in particular, as long as $c(- \mid \mathcal{L})$ is close to $u_D(- \mid \mathcal{L})$—then conditionalizing your priors on the evidence will approximate the result of conditionalizing u_D on the evidence. To illustrate, consider a variant of our earlier example (Fig. 5). This time suppose your priors are highly reliable, but not maximally so. You think that visual appearances are slightly more reliable than

they in fact are: $c(r \mid l \wedge ar) = .91$, whereas $u_D(r \mid l \wedge ar) = .9$. You now look outside and see that it appears to be raining. The most reliable evidentially constant updating process is not to conditionalize your actual priors on the strongest proposition you learn, but rather to conditionalize u_D on this proposition, and so to adopt a .9 posterior in r. Still, conditionalizing your priors on the evidence would *approximate* the result of u_D-conditionalization (since $.9 \approx .91$).

What if your priors are highly unreliable? Then the most reliable evidentially constant updating process is still to conditionalize u_D on your evidence. The greater the distance between your priors and u_D, the greater the distance between the process that maximizes reliability and the process that maximizes expected accuracy by your own lights. Recall irrational Ira, who has counterinductive priors. As before, let us suppose that Ira inhabits a world like ours, and so all of the relevantly similar situations are induction-friendly. If Ira updates by conditionalizing his actual priors on the evidence, then the more green emeralds he sees, the less confident he will be that the next unobserved emerald is green. While he will expect this to be the most reliable way of updating, he will be wrong about this: this updating process is highly unreliable, given his environment. He would fare better if he instead conditionalized an inductive prior on his evidence.

5.7 Reliably updating in non-partitional learning situations

Second, our reliabilist argument for Conditionalization was restricted to factive and partitional learning situations. The factivity assumption is relatively innocuous: this just means that your evidence consists only in truths—something that veritists should find independently attractive. The partitionality assumption is more controversial. It requires that the learning situation is transparent to the agent, in the sense that when they learn p, they also learn that they learn p. Externalists about evidence will be skeptical about whether evidence is always transparent.

What is the most reliable way of updating in opaque (non-partitional) learning situations? Let me mention two possibilities.

Schoenfield 2017 shows that the evidentially constant updating process that maximizes expected accuracy in all factive learning situations is *metaconditionalization*, which says to conditionalize not on the strongest proposition p that the agent learned, but rather on the proposition p': *the strongest proposition the agent learned is* p. By combining Schoenfield's result with the Reliability-Expected Accuracy Bridge for Updates, we can establish a more general conclusion. For any factive learning situation, if an agent's prior c achieves the highest possible reliability score, the most reliable evidentially constant updating process will metaconditionalize c on $\mathcal{L}$ (which will be the same as conditionalizing c on $\mathcal{L}$ in partitional learning situations).

Some epistemologists will recoil from the advice to metaconditionalize in opaque learning situations. After all, this is tantamount to advising agents to become certain of some proposition p' that they have not learned, and might have no reason to suspect is true (Schoenfield 2017, Gallow 2021; Meehan and Zhang 2025). This brings us to the second possibility, which is to opt for a stronger constraint on availability that rules out meta-

conditionalization. For example, Meehan and Zhang 2025 propose replacing evidential constancy with a stronger constraint on availability, *global evidential constancy*, according to which, roughly, if p is the strongest proposition the agent learns, then the credence in q produced by the updating process does not depend on the state at which the agent learned p or what they could have learned other than p. They show conditionalization satisfies this stronger constraint, whereas metaconditionalization does not.

Meehan and Zhang also show that, of the updating processes that satisfy global evidential constancy, the updating process with the highest total expected accuracy across all factive learning situations is conditionalization. Given the Reliability-Expected Accuracy Bridge for Updates, it follows that if an agent's prior achieves the highest possible reliability score, then most reliable globally evidentially constant updating process across factive learning situations is conditionalization. This would allow for a reliabilist defense of Conditionalization even in non-partitional settings.

For the purposes of this article, we do not need to settle this issue. The important point is that justification depends on whether an agent's posterior credence was generated by a sufficiently reliable updating process, where the threshold for sufficient reliability depends on what processes were genuinely available to them. While any plausible conception of availability should satisfy evidential constancy, there remains room for reasonable disagreement over whether a stronger constraint should be added and–if so—exactly what form it should take.

6 Headaches for reliabilists and the Bayesian cure

I have laid out a theory of justified credence that synthesizes Bayesianism and reliabilism. So far I have focused on the advantages this synthesis provides to Bayesians. But reliabilists also stand to benefit. This section considers three problems for classic reliabilist views. I show how the synthesis developed here solves each of them.

6.1 The problem of justified credence

Reliabilists typically focus on full belief. However, a complete epistemology should say something about both.[34]

The framework developed here fills this lacuna. On the view that emerges:

Justified Credence An agent's credence is justified iff it results from updating a justified prior credence in a rationally permissible manner.

[34] As noted in fn.3, there have been some important recent proposals for extending reliabilism to credences (Dunn 2016; Tang 2016; Pettigrew 2021). However, as noted there, none of these authors discuss how their proposals relates to Bayesianism. Since Bayesianism is the leading theory of rational credence in formal epistemology, I take it to be an advantage of the framework developed here that that it integrates the two. The proposals these authors put forward also lack some of the specific advantages of the framework developed here, such as the solution to the bootstrapping problem (to be discussed below).

Our reliabilist constraint on the priors tells us the conditions under which a prior is justified. And our reliabilist constraint on the posteriors tells us the conditions under which some update is rationally permissible.

6.2 Defeat

Reliabilist views struggle to account for defeat (Goldman 1979; Fumerton 1988; Lyons 2009; Lasonen-Aarnio 2010; Beddor 2015, 2021). A stock example:

> **Seeing Blue** Art is visiting an art exhibit. He sees a what appears to be a blue sculpture, and forms the belief that the sculpture is blue (b). Shortly thereafter, the curator introduces herself to Art, and explains that he is actually looking at a white sculpture illuminated by cleverly disguised blue lights. Art ignores the curator's testimony, retaining his conviction that the sculpture is blue. As a matter of fact, the curator had confused this sculpture with a different one. Art really was looking at a blue sculpture.

The curator's testimony defeats Art's justification for believing b. Classic externalist theories have trouble accommodating this intuition. For example, process reliabilism says that Art's belief is justified if it was formed through a reliable process. A natural way of typing his belief-forming process is *vision*. Suppose Art's eyesight is excellent, and the lighting conditions are normal. Then process reliabilism wrongly predicts that Art is still justified in believing b even after talking to the curator.

The problem extends beyond reliabilism. Art's belief may possess other externalist goodies. It could be safe (true in all nearby worlds where it is held), sensitive (if the sculpture were not blue, he would not have believed it), and apt (accurate in virtue of the exercise of a belief-forming competence). So Seeing Blue is a counterexample to any externalist view that says these properties suffice for justification.

Externalists have long acknowledged that their views face problems handling defeat (Goldman 1979: 101-103). Various fixes have been proposed, but many are plagued by further difficulties.[35] Rather than wading through all of the proposed fixes and further counterexamples, I want to show how my framework offers a natural solution.

Bayesians will think about this case in terms of Art's credences. Let c be Art's prior when he first sees the sculpture. Let e be Art's total evidence after speaking to the curator, and let c_e be his posterior after receiving this evidence. Since Art believes b before talking to the curator, $c(b)$ is fairly high. But what is $c(b \mid e)$? That is, what is Art's prior credence

[35] Goldman 1979 proposed to handle the problem by adding an additional clause involving 'alternative reliable processes'. The idea is that your belief is defeated as long as there is some alternative reliable process available to you which is such that, had you used it, you would no longer have held the belief in question. (See Grundmann 2009 and Bedke 2010 for related proposals.) However, this account has trouble handling more complex cases of defeater defeat (Lyons 2009, Beddor 2021); it also commits the conditional fallacy, rendering it susceptible to counterexamples (Beddor 2015); and it arguably abandons some of the reductive ambitions that rendered reliabilism attractive in the first place (Fumerton 1988).

that the sculpture is blue, conditional on being told by a curator that it is a white sculpture illuminated by a blue light?

The description of the case does not tell us. However, it's natural to assume that Art is aware that curators are knowledgeable about their exhibits, and that Art has no special reason to distrust this particular curator. Given these assumptions, it is natural to assume $c(b \mid e)$ is fairly low.

But it is also part of the description of the case that Art continues to believe that the sculpture is blue even after talking to the curator. Given the assumption that belief requires high credence, it follows that Art's posterior credence in r is high. So we have:

1. $c(b \mid e)$ = low
2. $c_e(b)$ = high

So Art is not updating by conditionalizing on his evidence.

This offers a diagnosis of why Art's posterior credences are not justified. Moreover, this is not just an internalist intrusion into an externalist theory. We have now given a reliabilist rationale for conditionalization. §5 showed that in factive and partitional learning situations, non-conditionalizers are guaranteed to fall shy of the reliabilist ideal. In Seeing Blue, it's natural to assume that Art's learning situation is factive, since he learns p: *the curator told me that this is a white sculpture illuminated by blue lights*, and p is true. It is also natural to assume it is partitional: he not only learns p; he also learns that he learns p. Given that he does not update by conditionalization, we can conclude that either (i) he did not update using the most reliable process available to him, or (ii) his priors were less than fully reliable. So our framework allows us to say that Art is not just exhibiting an internalist defect. He is also falling short of the reliabilist ideal.[36]

6.3 Bootstrapping troubles

Another influential objection is that reliabilism licenses a pernicious form of bootstrapping. Vogel 2000 presses this worry with examples like the following:

> **Roxy's Gas Gauge** Roxanne drives a car with a reliable gas gauge. She has never looked into the status of the gauge, or others like it. But she convinces herself the gauge is reliable using the following strange procedure. On Monday she looks at the gas gauge, which reads 'Full'. She forms the belief:
>
> > **Monday Reading** (R_M): The gauge reads 'full' on Monday.
>
> From which she infers:

[36] Of course, we saw earlier that ordinary people might not always be genuinely able to update by conditionalization, perhaps because of limited memory or computational abilities. But even in such cases, ordinary agents can consult their total evidence and adjust their credences in light of this evidence using some process that roughly approximates conditionalization. Yet this is not what Art does in Seeing Blue. He does not adjust his credences *at all* in light of the curator's testimony; he simply ignores it.

Monday Tank (T_M): The tank is full on Monday.

She conjoins these beliefs ($R_M \wedge T_M$), from which she infers:

Monday Accuracy (A_M): The gas gauge is accurate on Monday.

She repeats this process throughout the week, on each day forming the belief that the gauge is accurate on that day. On Saturday, she concludes by induction that:

Rel: The gas gauge is reliable.

There seems to be something epistemically defective about Roxanne's reasoning. But, Vogel argues, reliabilists have trouble explaining why. Since Roxanne's vision is reliable, reliabilism predicts that her beliefs about the gas gauge readings ('Reading beliefs') are justified. Since the gas gauge is reliable, reliabilism also seems to predict her beliefs about the level of gas in the tank ('Tank beliefs') are also justified. And since induction is reliable, it seems the reliabilist is forced to accept that she is justified in believing Rel on this basis.

Where does Roxanne's reasoning go wrong?

Weisberg 2010 offers a diagnosis. Weisberg notes that Roxanne infers *Rel* from her beliefs about the accuracy of the gauge, which are inferred from the conjunction of her Tank beliefs with her Reading beliefs. But her Tank beliefs are inferred solely from her Reading beliefs, which do not, taken on their own, provide the same level of support for *Rel*. Weisberg suggests that in all cases with this sort of structure, the agent's reasoning is defeated. Specifically, he proposes the following defeater on inductive reasoning:

No Feedback If (i) C is inferred from $L_1 \ldots L_n$ (and possibly other premises) by an argument whose justificatory power depends on making C at least x probable, (ii) $L_1 \ldots L_n$ are inferred from $P_1 \ldots P_n$, and (iii) $P_1 \ldots P_n$ do not make C at least x probable without the help of $L_1 \ldots L_n$, then the argument for C is defeated.

This diagnosis seems plausible, as far as it goes. But it raises some important questions. Is No Feedback a *sui generis* epistemic principle? Or can it be derived from some more general epistemological framework? If so, is this more general framework consistent with reliabilism?

A little reflection shows that No Feedback can be derived from the more general requirement to conditionalize on one's total evidence. Let c_{Sun} be Roxanne's prior credence before embarking on her bootstrapping procedure. Suppose that on each day of the week, the relevant evidence she acquires is just the Reading belief (e.g., R_M on Monday). On Sunday Roxanne has some prior credence that the gauge is reliable (Rel) conditional on all these Reading beliefs—i.e., $c_{Sun}(Rel \mid R_M \wedge R_T \wedge \cdots \wedge R_F)$. If she updates by conditionalization, her credence in *Rel* at the end of week is her old conditional credence:

$$c_{Sat}(Rel) = c_{Sun}(Rel \mid R_M \wedge R_T \wedge \cdots \wedge R_F) \tag{14}$$

So if she updates by conditionalization, she doesn't get to use her Tank beliefs to 'boost' her credence in Rel beyond what was supported by the original Reading beliefs.

Consequently, Weisberg's No Feedback condition can be derived from the requirement to update by conditionalization. This gives us a simple and principled diagnosis of Roxanne's error. Moreover, it is also a diagnosis that is compatible with reliabilism, in view of our argument in §5. Roxanne, much like Art, falls short of the reliabilist ideal.

7 Concluding Remarks

Reliabilism and Bayesianism have usually been developed in isolation from one another. This paper advocated for a rapprochement. I argued that the internalist orientation of traditional Bayesian approaches is holding Bayesians back from reaching their full epistemological potential. Importing some ideas from reliabilist epistemology provides a new solution to the problem of the priors. A reliabilist perspective also furnishes new, externalist arguments for the core Bayesian norms (Probabilism and Conditionalization).

The resulting synthesis also carries benefits for reliabilists. Reliabilists face longstanding problems accounting for justified credence, defeat, and bootstrapping. This proposal advanced here offers a solution. I have given a reliabilist treatment of justified credence and shown how it can be integrated with the leading theory of rational credence in formal epistemology (Bayesianism). And the theory offers a principled way of handling cases of defeat and bootstrapping. All of these cases involve agents who fail to update by conditionalization, and hence fail to use the most reliable available updating process.

References

William Alston. An internalist externalism. *Synthese*, 74(3):265–283, 1988.

D.M. Armstrong. *Belief, Truth, and Knowledge*. Cambridge University Press, London, 1973.

Christina Ballarini. Epistemic blame and the new evil demon problem. *Philosophical Studies*, 179(8):2475–2505, 2022.

Bob Beddor. Process reliabilism's troubles with defeat. *The Philosophical Quarterly*, 65 (259), 2015.

Bob Beddor. Reasons for reliabilism. In Brown and Simion, editors, *Reasons, Justification, and Defeat*. Oxford University Press, 2021.

Bob Beddor and Carlotta Pavese. Modal Virtue Epistemology. *Philosophy and Phenomenological Research*, 101(1):61–79, 2020.

Matthew Bedke. Developmental process reliabilism: on justification, defeat, and evidence. *Erkenntnis*, 73(1):1–17, 2010.

Jakob Bernoulli. *Ars Conjectandi*. Impensis Turnisiorum Fratrum, Basel, 1713.

Sharon Berry. External world skepticism, confidence and psychologism about the priors. *The Southern Journal of Philosophy*, 57(3):324–346, 2019.

Glenn Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

Tyler Burge. Perceptual entitlement. *Philosophy and Phenomenological Research*, 67(3): 503–48, 2003.

Rudolf Carnap. A basic system of inductive logic, part i. In Richard C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, pages 34–165. University of California Press, 1971.

Stewart Cohen. Justification and Truth. *Philosophical Studies*, 46:279–296, 1984.

Nilanjan Das. Externalism and exploitability. *Philosophy and Phenomenological Research*, 104(1):101–128, 2022.

Bruno de Finetti. *Theory of Probability*, volume 1. Wiley, New York, 1974.

Igor Douven. Inference to the best explanation, dutch books, and inaccuracy minimization. *The Philosophical Quarterly*, 63(252):428–444, 2013.

Jeffrey Dunn. Reliability for degrees of belief. *Philosophical Studies*, 172:1929–1952, 2016.

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

Benjamin Eva. Principles of indifference. *Journal of Philosophy*, 116(7):390–411, 2019.

Roderick Firth. Are epistemic concepts reducible to ethical concepts? In Alvin Goldman and Jaegwon Kim, editors, *Values and Morals*. D. Reidel, Dordrecht, 1978.

Richard Fumerton. Foundationalism, conceptual regress, and reliabilism. *Analysis*, 48(4): 178–184, 1988.

Dmitri Gallow. Updating for externalists. *Noûs*, 55(3):487–516, 2021.

Alvin Goldman. What is justified belief? In Pappas, editor, *Justification and Knowledge*. Reidel, Dordrecht, 1979.

Alvin Goldman. *Epistemology and Cognition*. Harvard University Press, Cambridge, MA, 1986.

Alvin Goldman. *Knowledge in a Social World*. Oxford University Press, New York, 1999.

Jeremy Goodman and Bernhard Salow. Taking a Chance on KK. *Philosophical Studies*, 175: 183–196, 2018.

Nelson Goodman. *Fact, Fiction, and Forecast*. Harvard University Press, Cambridge, 1955.

Peter J Graham. Epistemic Entitlement. *Noûs*, 46(3):449–482, 2012.

Peter J Graham. Normal circumstances reliabilism. *Philosophical Topics*, 45(1):33–61, 2017.

Peter J Graham. The New Evil Demon Problem at 40. *Philosophy and Phenomenological Research*, forthcoming.

Hilary Greaves and David Wallace. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632, 2006.

Daniel Greco. Justification and Excuses in Epistemology. *Noûs*, 55(3):517–537, 2021.

Thomas Grundmann. Reliabilism and the problem of defeaters. *Grazer Philosophische Studien*, 79(1):65–76, 2009.

Ned Hall. Correcting the guide to objective chance. *Mind*, 103(412):505–518, 1994.

Sophie Horowitz. Immoderately rational. *Philosophical Studies*, 167(1):1–16, 2013.

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

Christoph Kelp and Mona Simion. Criticism and Blame in Action and Assertion. *Journal of Philosophy*, 114(2):76–93, 2017.

John Maynard Keynes. *A Treatise on Probability*. Macmillan and Co., London, 1921.

Pierre-Simon LaPlace. *Essai Philosophique sur les Probabilités*. Courcier, 1814.

Maria Lasonen-Aarnio. Unreasonable Knowledge. *Philosophical Perspectives*, 24:1–21, 2010.

Maria Lasonen-Aarnio. *The Good, the Bad, and the Feasible: Knowledge and Reasonable Belief*. Oxford University Press, Oxford, forthcoming.

Hannes Leitgeb and Richard Pettigrew. An objective justification of bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77(2):236–272, 2010.

Jarrett Leplin. In Defense of Reliabilism. *Philosophical Studies*, 134(1):31–42, 2012.

David Lewis. A Subjectivist's Guide to Objective Chance. In Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume 2, pages 263–293. University of California Press, 1980.

David Lewis. Humean supervenience debugged. *Mind*, 103(412):473–490, 1994.

Clayton Littlejohn. A Plea for Epistemic Excuses. In Dorsch and Dutant, editors, *The New Evil Demon*. Oxford University Press, Oxford, forthcoming.

Sebastian Liu. How to be indifferent. *Noûs*, 59(2):317–334, 2025.

Jack Lyons. *Perception and Basic Beliefs*. Oxford University Press, New York, 2009.

Pablo Zendejas Medina. Just as planned: Bayesianism, externalism, and plan coherence. *Philosophers' Imprint*, 2023.

Alexander Meehan and Snow Zhang. Bayes is back. *The Philosophical Review*, 134(4): 285–350, 2025.

Graham Oddie. Conditionalization, Cogency, and Cognitive Value. *British Journal for the Philosophy of Science*, 48:533–541, 1997.

Richard Pettigrew. Accuracy, risk, and the principle of indifference. *Philosophy and Phenomenological Research*, 92(1):35–59, 2016a.

Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, Oxford, 2016b.

Richard Pettigrew. What is justified credence? *Episteme*, 18(1):16–30, 2021.

Joel Predd, Robert Seiringer, Elliott Lieb, Daniel Osherson, Vincent Poor, and Sanjeev Kulkani. Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792, 2009.

Leonard Savage. *The Foundations of Statistics*. Dover Publications, New York, 2 edition, 1972.

Miriam Schoenfield. Conditionalization does not (in general) maximize expected accuracy. *Mind*, 126(504):1155–1187, 2017.

Abner Shimony, R. Sherman Lehman, and John G. Kemeny. Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 33(3):481–482, 1968.

Paul Silva and Luis R.G. Oliveira. Propositional justification and doxastic justification. In Lasonen-Aarnio and Littlejohn, editors, *The Routledge Handbook of the Philosophy of Evidence*. Routledge, 2023.

Martin Smith. *Between Probability and Certainty: What Justifies Belief*. Oxford University Press, Oxford, 2016.

Robert Smithson. The principle of indifference and inductive skepticism. *British Journal for the Philosophy of Science*, 68(1):253–272, 2017.

Weng Hong Tang. Reliability theories of justified credence. *Mind*, 125(497):63–94, 2016.

Bas van Fraassen. *Laws and Symmetry*. Oxford University Press, Oxford, 1989.

Jonathan Vogel. Reliabilism leveled. *Journal of Philosophy*, 97:602–623, 2000.

Jonathan Weisberg. Bootstrapping in general. *Philosophy and Phenomenological Research*, 81(3):525–548, 2010.

Roger White. Evidential symmetry and mushy credence. *Oxford Studies in Epistemology*, 3:161–186, 2009.

Michael Williams. Internalism, reliabilism, and deontology. In Hilary Kornblith and Brian McLaughlin, editors, *Goldman and his Critics*, pages 1–21. Blackwell, 2016.

Jon Williamson. Justifying the principle of indifference. *European Journal for Philosophy of Science*, 8(3):559–586, 2018.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

Timothy Williamson. Justifications, Excuses, and Sceptical Scenarios. In Dorsch and Dutant, editors, *The New Evil Demon Problem*. Oxford University Press, Oxford, forthcoming.