

# The reliability problem for reliabilism

Matthew Frise<sup>1</sup>

Published online: 28 March 2017  
© Springer Science+Business Media Dordrecht 2017

**Abstract** According to process reliabilism, a belief produced by a reliable belief-forming process is justified. I introduce problems for this theory on any account of reliability. Does the performance of a process in some domain of worlds settle its reliability? The theories that answer “Yes” typically fail to state the temporal parameters of this performance. I argue that any theory paired with any plausible parameters has implausible implications. The theories that answer “No,” I argue, thereby lack essential support and exacerbate familiar problems. There are new reasons to avoid any reliability conditions on justification.

**Keywords** Process reliabilism · Reliabilism · Reliability problem · Temporality problem · Alvin Goldman

## 1 Introduction

Typical beliefs based on perceptual experience seem justified. Guesses don’t. Any theory of justification that respects these intuitive assessments has some support. Here is a simplified statement of Goldman’s (1979) process reliabilism, a widely-held theory that seems to respect these assessments: S’s belief that  $p$  is justified *iff* S’s belief that  $p$  is the output of a reliable belief-forming process. Process reliabilism (hereafter *reliabilism*) implies that any belief that is reliably formed is justified. A reliable process is one that tends to produce true rather than false beliefs. Beliefs formed by perceptual experience tend to be true, and according to reliabilism this explains why they are justified. Guesses tend to be false. That

---

✉ Matthew Frise  
mjfrise@gmail.com

<sup>1</sup> Department of Philosophy, Baylor University, One Bear Place #97273, Waco, TX 76798-7273, USA

explains why they aren't justified, according to reliabilism. Reliabilism appears to have some support.

As a result, epistemologists now commonly celebrate the reliable and denigrate the unreliable. This is particularly so among externalists about epistemic justification. Many externalists who aren't reliabilists incorporate reliability conditions in their theory of justification.<sup>1</sup> Externalism is the view that mentally alike subjects can differ in their justification. Some mentally identical subjects might not *justifiedly* believe the same things. This may be because, say, the processes forming these beliefs are reliable in one subject's environment, but not in the other's.

Reliabilists have not filled in key details of the reliability that they say matters for justification, and no one has encouraged this task. Most reliabilists think reliability is a matter of performing well enough. But the temporal measure of this performance is unclear. Other reliabilists deny that performance determines reliability. But then it is unclear how to test for reliability. I call the *reliability problem* the problem of supplying a suitable account of reliability. I argue that any account introduces new troubles for reliabilism. I begin with a case study, showing in Sect. 2 how a version of reliabilism faces the reliability problem, and distinguishing this problem from reliabilism's generality problem. In Sect. 3 I argue that the best solutions are inadequate. In Sect. 4 I argue that the same holds for every defended version of reliabilism. There is new reason to avoid externalist theories that involve reliability in necessary or sufficient conditions for justification.

## 2 The reliability problem for actual world reliabilism

Reliability is a tendency to produce true rather than false beliefs. But what exactly is this tendency? When Goldman (1979: 11) first used the term, he said it could refer "either to the *actual* long-run frequency, or to a 'propensity', outcomes that would occur in merely *possible* realizations of the process." Goldman (2009: 252) fleshes out this first option:

Reliability of a process should first be judged by the truth ratio of its belief outputs in the actual world, and then this reliability score should be rigidified. Rigidification implies that [for instance] visually-based beliefs produced by the same process in a demon-world will be equally justified as similarly produced beliefs in the actual world.<sup>2</sup>

A reliable process of belief-formation produces a sufficiently high ratio of true beliefs here in the actual world. A belief formed in any world by a process that is reliable in ours is justified. In other words:

<sup>1</sup> See e.g., Bergmann's (2006: 133) proper functionalism, Greco's (1999: 286–289) agent reliabilism, and Sosa's (2001) virtue reliabilism.

<sup>2</sup> Cf. Goldman (2008a).

*Actual World Reliabilism* (AWR). S's belief that  $p$  is justified *iff* S's belief that  $p$  is the output of a belief-formation process that is reliable in the *actual* world.<sup>3</sup>

AWR is an interpretation of a schema I call:

*Lone World Reliabilism*. S's belief that  $p$  is justified *iff* S's belief that  $p$  is the output of a belief-formation process type that is reliable in  $w$ .

Lone World Reliabilism states that a justifying belief-formation process is reliable in a single-membered domain of worlds,  $w$ . AWR identifies  $w$  as the actual world. The justificatory power of a belief-formation process is determined just by how accurate the process actually is. And we have a fair amount of information about how things actually are. We needn't speculate about less familiar worlds. Because AWR is so straightforward, it is helpful to begin with how it in particular faces the reliability problem. Later, in Sect. 4, I will present the problem for all other defended versions of reliabilism, including those that understand *tendency* as a propensity.

Actual World Reliabilism (AWR) identifies the domain of worlds in which a process's performance matters. But for any identified domain, which outputs in it matter at a given time? Some processes have produced beliefs for thousands of years, and will continue to produce them for thousands more. Do beliefs that haven't been formed yet, and beliefs from the distant past, affect the truth ratio that matters now? Reliabilists generally don't say.

Even Goldman is ambiguous on the matter. He (1979: 17) imagines a scenario in which "wishful thinking turns out to be reliable *in the actual world*" because "there is a benevolent demon who, lazy until now, will shortly start arranging things so that our wishes come true. The long-run performance of wishful thinking will be very good..." Goldman thinks the truth-values of some outputs that will be produced shortly are among those that affect the process's truth ratio that matters. But do they *all* affect it? Do they affect it *shortly*, or do they also affect it now, and in the distant past? And do the truth-values of distant-future and distant-past outputs affect it? It's unclear just which of a process's outputs determine its "long-run performance".

Yet these sorts of details matter—they give AWR importantly different implications about justification. Most processes produce true beliefs in some cases and false beliefs in others. The truth ratio of the beliefs formed in one span of time often importantly differs from the truth ratio of beliefs formed in another span. Storm forecasting from observations is accurate more often now that we have certain observation technology. If climate change makes our weather more erratic,

<sup>3</sup> AWR descends from a theory of the truth-conditions for *statements* about justification (cf. Comesaña 2002; Goldman 1992). For discussion see Becker (2007), Goldman (1986), and Lyons (2013). Goldman (2009) strongly favors AWR. And Goldman (2008a) seems to defend AWR—he discusses several versions of reliabilism, and leaves all but AWR with unanswered objections. When asked about AWR, Goldman said in conversation that he changes his mind about which reliabilist view is right.

Throughout, for ease of exposition I simplify reliabilism, omitting details about belief-dependent belief-forming processes, belief-sustaining processes, and *prima facie* justification. Nothing hangs on this.

storm forecasting may, despite our technology, become accurate less often. Given these changes, it is not obvious whether this forecasting counted, currently counts, and will count, as reliable. AWR needs an account of the temporal parameters during which a process's truth ratio determines its reliability at a time. Without this account, AWR has no implications about the reliability of storm forecasting or of most other processes.<sup>4</sup> So, without this account, AWR is silent about the justificatory status of most beliefs. And so the reliability problem manifests itself here as what I call the *temporality problem*, the problem of providing a principled, explanatory account of the temporal parameters which settle a process's reliability at a time, and thus its justificatory power at a time. Since any version of reliabilism is supposed to explain the justificatory status of any belief, finding a suitable account is critical to AWR's survival. And finding it is not easy.

With few exceptions reliabilists have not proposed, much less defended, any accounts. There is an unlimited number. So that we can assess AWR's implications, I'll propose five accounts that are at least as plausible as any. We could complicate things by hybridizing these accounts, by adding further temporal parameters, or by focusing on subsets of beliefs within the temporal parameters. I will keep things simpler. On the first three accounts, reliability is what I call *static*. If reliability is static, a process is either reliable at all times or at none—its reliability does not change with time.

The first static account is:

*Enduring Reliability.* A process of belief-formation  $r$  is reliable *iff* for any time  $t$  such that  $r$  has formed a belief by  $t$ , the truth ratio of all beliefs  $r$  has formed by  $t$  is good enough.

Since the reliabilist view we are presently supplementing is AWR, we are only interested in whether  $r$  is reliable in our world. So, in this section and the next, understand the left-hand side and right-hand side of any account of reliability to be indexed to our world. Since Enduring Reliability is a static account, the left-hand side needn't relativize reliability temporally. For simplicity I will omit time variables on the left-hand side in static accounts. A process has a truth ratio that is *good enough* for reliability if this ratio is at or beyond a certain minimum. Goldman (1986: 103) says this minimum is over 1-to-1, but he does not say how far over. Later he (1986: 105) says, more weakly, it is probably not less than 1-to-1. For simplicity, I'll assume it is slightly over 1-to-1. We could ratchet the ratio up, or make it vague, and the problems I introduce below still arise.

Enduring Reliability says that a process is reliable just in case it has, from its first tokening onward, always produced more true beliefs than false beliefs. This account looks at all future performance of the process. It is also a *track-record* account, looking at all of the past performance too. And the account is *complex*, looking at the performance during multiple intervals of time; in particular, it looks at the performance during every interval that starts with the process's first use. Enduring

<sup>4</sup> Cf. Baumann (2009: 85–86). Why “most”? Because some processes only ever produce true beliefs, and others only ever produce false beliefs. Whether these processes are reliable is independent of temporal parameters.

Reliability is fairly demanding. If we pair Enduring Reliability with AWR, a justification-conferring process in any world is one with a truth ratio that *has* never dipped, and *will* never dip, beneath a certain threshold in the actual world. This captures the idea that a justifying process is one that at no point in fact has a bad track-record.

The next static account is:

*Opportune Reliability.* A process of belief-formation  $r$  is reliable *iff* the truth ratio of all beliefs  $r$  forms by  $t_n$  is good enough.

On this account, a process is reliable just in case at a certain important time ( $t_n$ ) its overall truth ratio is good enough. Opportune Reliability is a track-record account like Enduring Reliability. But Opportune Reliability is *simple* rather than complex, looking at a process's performance during just one interval of time. It requires just an adequate overall performance by  $t_n$ , while Enduring Reliability requires an adequate overall performance from the start and ever after. On one fleshing out,  $t_n$  is fairly soon, and so Opportune Reliability allows a process to count as reliable even if its truth ratio plummets in the distant future. On another fleshing out,  $t_n$  is the time at which a process produces its last output ever, and so its reliability at any time is determined by the truth ratio of its most inclusive span of outputs.<sup>5</sup> This fleshing out resembles Enduring Reliability, but is weaker and perhaps more intuitive. It requires only a good enough performance ultimately, in the end, not one that endures to the end. On Opportune Reliability, AWR implies that a justifying process in any world is one that has a good enough track-record at some special time in the actual world.

The final static account is:

*Period Reliability.* A process of belief-formation  $r$  is reliable *iff* the truth ratio of all beliefs that  $r$  forms between  $t_m$  and  $t_n$  is good enough.

Here, a process is reliable just in case the truth ratio of all beliefs it produces during a certain important period ( $t_m$  through  $t_n$ ) is good enough. Period Reliability is simple, promoting the notion that a reliable process is one that performs satisfactorily during an important, fixed window of time, even if not before or after. Since  $t_m$  may be sometime after a process first produces a belief, Period Reliability isn't a track-record account. As a result, a process that has an especially rough start could still count as reliable on Period Reliability, but might not on Opportune Reliability (and *would* not on Enduring Reliability). If Period Reliability identifies the kind of reliability that matters for justification, then justifying processes in any world, according to AWR, do well enough in the important window in the actual world.

On the remaining accounts that I consider, reliability is what I call *dynamic*. It is possible that a process is reliable at one time but not at another. This raises a question. How exactly do dynamics in reliability trickle down to justification? That

<sup>5</sup> Goldman (1986: 49) is ambiguous between this version of Opportune Reliability and Historical Reliability (see below): "We may take [reliability] to refer to the truth ratio of beliefs generated by the process on all actual occasions of use". Does he mean all actual occasions *there will ever be*, or all actual occasions *to date*?

is, how, if at all, does the justification of a belief vary when the reliability of its formation process varies over time? Suppose that, at noon today, reliable process  $r$  forms Jones' belief that all stinky cheeses stink. If at midnight  $r$  becomes unreliable, is it true at midnight that Jones' belief at noon was justified? If so, is it still justified at midnight?

One trickle-down theory is that a belief is justified *exactly while* its process of formation is reliable. That is, if process  $r$  forms a belief that  $p$  at  $t_1$ , the belief that  $p$  is justified at  $t_2$  just if  $r$  is reliable at  $t_2$ . So, Jones' belief is justified until, but not at, midnight. Another trickle-down theory is that a belief is justified just in case its process of formation is reliable *at the time of its formation*. That is, if process  $r$  forms a belief that  $p$  at  $t_1$ , the belief that  $p$  is justified at any time just if  $r$  is reliable at  $t_1$ . Jones' belief remains justified at midnight. There are indefinitely many ways to map justification on to beliefs formed by dynamically reliable processes. I won't state them all. I will employ the first view, since the others in the end aid AWR no better: justification co-varies with reliability.

The first dynamic account is:

*Historical Reliability.* A process of belief-formation  $r$  is reliable at time  $t$  iff the truth ratio of all beliefs  $r$  forms by  $t$  is good enough.

I include a time variable in the left-hand side of the account in order to emphasize its dynamic character. A process is reliable at  $t$  just when most of its outputs formed *to date* (up through  $t$ ) are true. This track-record account allows reliability to vary whenever a change in environment, cognitive behavior, luck, etc., notably shifts a process's overall performance. If a process was reliable at noon last New Year's Eve, it had performed adequately by then; if unreliable at noon next New Year's Eve, it will have performed inadequately by then. A process's reliability at a time hinges on its complete history then. Unlike with static accounts, reliability here hinges on something that changes with time. Yet Historical Reliability is a simple account, since performance matters in only one temporal interval at a time. AWR, on Historical Reliability, states that a process confers justification at those times where its truth ratio so far in the actual world is high enough. Tomorrow's belief that  $p$  may have a different justificatory status from that of today's belief that  $p$ .<sup>6</sup>

The final account is:

*Proximate Reliability.* A process of belief-formation  $r$  is reliable at time  $t$  iff the truth ratio of all beliefs that  $r$  forms near  $t$  is good enough.<sup>7</sup>

Reliability is dynamic on this simple account. But unlike Historical Reliability, this is not a track-record account. Here, a process is reliable at a time just in case it mostly produces true beliefs near that time. Distant past and future outputs don't matter. As with Period Reliability, reliability depends on performance within a period of time, but here the relevant period changes. Since a process's truth ratio can

<sup>6</sup> It is difficult to identify times in non-actual worlds with times in the actual world. Yet AWR must do this given any dynamic account of reliability. In the interests of charity, however, I will not assume such an identification is impossible. I thank an anonymous reviewer for noting the difficulty.

<sup>7</sup> See Comesaña (2002: 259–260).

importantly change near different times, the reliability of a process can change too. Near noon last New Year's Eve, a process might perform poorly, but it might do better near noon next New Year's Eve. The process would be unreliable at the first noon and reliable at the second. AWR, on Proximate Reliability, allows the justification of a belief to vary over time with the changes in reliability of its formation process. Depending on how the future unfolds, a belief can have one justificatory status today and another tomorrow.

In the next section I will show why none of the above solutions to the temporality problem is good. First, I will make a point about reliabilism's generality problem. On reliabilism, the justification of any belief is determined by the reliability of the type of process that produced it. But any belief is formed by a token process of an unlimited number of types. The generality problem is that of specifying in a principled way the type relevant for determining the justificatory status of a given belief. And there appears to be no good way to do this.<sup>8</sup>

It has been suggested to me that the temporality problem does not ultimately differ from the generality problem.<sup>9</sup> I will more than suggest the suggestion lacks support. The main discussants of the generality problem have never understood that problem to concern the issue of clarifying the term *reliability*. They present those as distinct challenges for reliabilism.<sup>10</sup>

Maybe the suggestion arises because (1) some solution to one of these problems also solves the other. While (1) seems true, it does not indicate that the problems are the same—therapy may solve my gambling and drinking problems, but sadly they are distinct. Maybe it arises because (2) any solution to one of these problems also solves the other. But (2) is false.

A solution to the temporality problem is not thereby a solution to the generality problem. Enduring Reliability, for example, provides a solution to the temporality problem (whether it provides a *good* solution remains to be discussed). It tells us what determines the reliability of each process type that forms a given belief. But it does nothing to specify which of these types is *the* relevant one—which type determines that belief's justification. Of course, having information about which of the process types is reliable is crucial for *adequately* solving the generality problem. But the information fits countless solutions. So even if we had it, the hard work of finding an adequate solution remains. This just displays the seriousness of the temporality problem—solving it appears to be critical for solving what had seemed to be reliabilism's worst problem.

Also, a solution to the generality problem is not thereby a solution to the temporality problem. A solution to the generality problem can be understood as specifying the process type the reliability of which determines the justification of a belief formed in a given circumstance.<sup>11</sup> For any belief formed in circumstance  $C_1$ ,

<sup>8</sup> See Conee and Feldman (1998), Feldman (1985), Feldman and Conee (2002), and Pollock (1984).

<sup>9</sup> I thank several anonymous referees at other journals for this suggestion.

<sup>10</sup> See, e.g., Goldman (1986: 49–50, 1979: 12–17) and Feldman and Conee (1985: 25–26). Both challenges appear to be versions of the reference class problem; cf. Baumann (2009: 84–87). Still, the challenges are not identical.

<sup>11</sup> Cf. Alston (1995: 8) and Pollock (1984: 108).

the relevant process is forming belief via *visual experience*; for any belief formed in circumstance  $C_2$ , the relevant process is forming belief via *visual experience in bad lighting*; complete this for all  $C_i$ , and state exactly what  $C_1$ ,  $C_2$ , etc. are. But this does not solve the temporality problem. Suppose belief is formed in  $C_1$  at  $t$ . At  $t$ , which outputs determine the reliability of *visual experience*? It will be the outputs formed in *some* instances of  $C_1$ . But which? All instances of  $C_1$  that obtain before  $t$ ? Or also all future instances of  $C_1$ ? Or maybe just all instances of  $C_1$  near  $t$ ? The temporality problem remains. Still, one might think the *best* solution to the generality problem will solve the temporality problem too. But even if this is correct, it does not reveal that the problems are identical. Therapy is the best solution to plenty of distinct problems.

And even if the temporality problem ultimately just is that intractable, familiar-as-Friday generality problem, then this paper shows that the generality problem is more severe than we realized. A solution to the generality problem will have to individuate relevant process types at a finer grain than we thought: at the level of the temporal window in which all their tokens occur. A solution might state that, for example, in any instance of  $C_1$  that obtains this year, the relevant process type is *visual experience between the years 2000 and 2034*. This isn't how we are used to typing processes, so the relevant ones will be unfamiliar and therefore harder to identify, and reliabilism is less commonsensical than reliabilists have led us to believe.<sup>12</sup> More importantly, if my points below hold, then no solution that individuates at this level of granularity is good. So the generality problem, for a new reason, looks unsolvable.

### 3 The severity of the temporality problem

I have stated the temporality problem and some accounts of reliability. There are particular problems for each account, preventing each from solving the temporality problem. Rather than cover these, I will introduce three general problems showing that none of these accounts, or any other, adequately solves the temporality problem. This is bad for AWR. It lacks the right implications about the justificatory status of nearly any belief.

#### 3.1 The unnatural deduction problem

The first new problem for AWR on each account of reliability requires assembly. I will build it in several steps. The first step reflects on some of our higher-level justification. We have some strongly justified beliefs about which beliefs are justified and which are unjustified. Let  $G$  = a belief formed by apparent guessing (call it a *guess*).  $G$  seems to the subject formed arbitrarily, seems formed without enough relevant information, without even a hunch or a memory of a relevant good

<sup>12</sup> Notably, reliabilists like Henderson and Horgan (2010: 91) think that to identify the relevant processes in this way is unscientific and “really is a convenient proxy for what ultimately presumably matters”.



track-record. Let B = a belief that takes perception at face value. We believe with justification that:

P1. G is unjustified.

P1\*. B is justified.

In order for AWR to be correct, it must accommodate our having justified beliefs in things like (P1) and (P1\*).<sup>13</sup> Presumably the relevant process that forms G is *apparent guessing* (hereafter, *guessing*). The relevant process that forms B is *taking perception at face value* (hereafter, *perception*), where perceiving is non-factive.<sup>14</sup> By understanding AWR, we can reasonably believe:

P2. If AWR is true and G is unjustified, then guessing is unreliable.

P2\*. If AWR is true and B is justified, then perception is reliable.

The various accounts of reliability identify various “windows” of time in which a process’s truth ratio determines its reliability. Call a process’s truth ratio in the temporal window (or windows, in the case of complex accounts like Enduring Reliability; I’ll use the singular for simplicity) its *window-truth ratio*. On any account of reliability, it’s the case that:

P3. If guessing is unreliable, then guessing’s window-truth ratio is not good enough.

P3\*. If perception is reliable, then perception’s window-truth ratio is good enough.

And here comes trouble. Over time our environment and our cognitive behavior in it change. The performance of our belief-forming processes to date changes too. Was the actual world ever *significantly* different, or will it ever be, such that guessing used to or will mostly yield true beliefs overall, or such that perception used to or will mostly yield false beliefs overall? Consider two stories:

*Blindluck.* System 1 perceptual processing actually developed in our evolutionary chain before System 2 perceptual processing. That is, the sensory systems of our predecessors took in information about their surroundings before it was consciously accessible to them. They had to form

<sup>13</sup> BonJour (2001/2010: 218–219) thinks externalist theories of justification can’t accommodate this because we lack *direct* access to the external factors that determine the justificatory status of beliefs like B or G. Kornblith (2004: 198) thinks externalism can accommodate. Yet BonJour and Kornblith agree that a theory of justification *should* accommodate our intuitions about our object-level justification. They are right. Epistemology seems hopeless otherwise. Epistemological theories seem meritless if they are selected to fit meritless intuitive judgments about justification. However, unlike BonJour, I am pointing out what follows if AWR *can* accommodate the justification of the judgments.

<sup>14</sup> One can evaluate reliabilism only by stipulating which process of belief-formation is relevant. For my purposes it ultimately does not matter which ones we stipulate. Comesaña (2002) and Sosa (2003) take *taking perception at face value* to be the relevant process that forms beliefs like B (perhaps this can be spelled out as forming beliefs that fit the representational contents of perceptual experience). Cf. Goldman (2008a, 1986, 1979), who moreover takes guessing to be the relevant process that forms beliefs like G. An adequate solution to the generality problem may individuate relevant processes at a finer grain than these.

beliefs about their environment by guessing, when making conscious decisions. But, as with familiar cases of blindsight, these beliefs incorporated undetectable environmental information from System 1 processing. As a result, most guesses were correct. Guessing in this way continued for some while. Eventually System 2 perceptual processing developed, and along with it, processes of forming beliefs about one's surrounding using consciously available information. Relatively few guesses have been formed since. To date, guessing has produced many more true beliefs than false beliefs, and that ratio won't change.

*Experience Machine.* In the actual future, most of Earth's many billions of people will live in a false reality presented by an Experience Machine. The machine simulates a desire-fulfilling reality for anyone hooked up to it, and erases from the subject's memory any indication that her experiences are misleading. People hooked up to the machine continue to form beliefs based on what they experience. Eventually, perception will have yielded many more false beliefs than true ones.<sup>15</sup>

If guessing is unreliable, then nothing like Blindluck obtains during the temporal window in which a process's truth ratio determines its reliability. If perception is reliable, then nothing like Experience Machine obtains during that window either. What's more, no scenario of the same general *type* as Blindluck or Experience Machine occurs then. The (long!) disjunction of every story in which guessing has mostly yielded true beliefs in the window would be false, and the (long!) disjunction of every story in which perception will mostly yield false beliefs in the window would be false:

P4. If guessing's window-truth ratio is not good enough, then no Blindluck-type scenario occurs in the window.

P4\*. If perception's window-truth ratio is good enough, then no Experience Machine-type scenario occurs in the window.

And the sequences of premises support:

C. If AWR is true, then no Blindluck-type scenario occurs in the window.

C\*. If AWR is true, then no Experience Machine-type scenario occurs in the window.

Each premise is supported, and we can appreciate their connections. So it's reasonable to believe (C) and (C\*). But if it's reasonable to believe (C), there are serious problems with Enduring Reliability, Opportune Reliability, and Historical

<sup>15</sup> Blindluck resembles an example that presents what I'll call the clairvoyance problem for reliabilism: some *possible* subject without any relevant information forms beliefs by a reliable process, and reliabilism implausibly counts these beliefs as justified (see BonJour 1980). Experience Machine resembles an example that presents the new evil demon problem for reliabilism: a massively deceived *possible* subject who is mentally just like me should also share all my justification, but reliabilism implausibly implies otherwise since just one of us reliably forms beliefs (see Cohen 1984). However, Blindluck and Experience Machine present new problems because they, unlike the examples, concern *actual* beliefs and concern the reliability of processes in the *actual* world, processes whose alleged actual tendencies have been thought to motivate reliabilism.

Reliability as solutions to the temporality problem. Being track-record accounts, they include all past activity of the process in the temporal window in which the reliability of a belief-forming process is determined. AWR, paired with any of these views, implies that no Blindluck-type scenario occurred at any relevant past time.

But we don't have much reason to believe this implication, the consequent in (C). This is not because we lack *direct* access to relevant external factors. It is because the implication is about an underdetermined empirical matter. We don't have nearly enough information about the past. It's unclear whether our predecessors guessed much more often or accurately than we do. There is substantial support for the *empirical* thesis that System 2 processing evolved after System 1. Since we cannot reasonably deny from the armchair the disjunction of all stories in which guessing has, to date, mostly yielded true beliefs, we cannot reasonably believe AWR when paired with any of these accounts of reliability.

Similarly, on Enduring Reliability and the more intuitive versions of Opportune Reliability, the future shapes the window-truth ratio of processes. AWR paired with either, given (C\*), implies that no Experience Machine-type scenario will occur. But it is unreasonable for us to believe that implication, and so these pairings are unreasonable. It's far from clear how our environment and technology will change, and how we will form beliefs in response, and with what success. Just consider how clear, 200 years ago, developments in these areas would have been relative to today! These considerations make it dubious simply to reason inductively that the future and distant past track-records of guessing and perception resemble their recent track-records.

Before I show how much worse the problem here gets, I will address some potential needling from the generality problem, which might seem to unstitch my claims. Apparent guessing is not identical to the process of forming belief by using undetectable environmental information. And perception is not identical to the process of forming belief based on experiences induced by an experience machine. How can I nonetheless uphold (P4) and (P4\*), and connect guessing and perception with the Blindluck and Experience Machine stories?

Here's how. Guessing and perception are general process types. The belief-formation processes in the stories are in fact just *subtypes* of guessing and perception (respectively). And if process  $r_1$  is a subtype of process  $r_2$ , then  $r_1$ 's outputs affect  $r_2$ 's truth ratio; any output of  $r_1$  is also an output of  $r_2$ . So, the performance of the subtypes mentioned in the stories affects the performance of guessing and perception. Now, the generality problem of course raises such questions as: if a belief is the output of guessing and of a subtype of guessing, which circumstances select which of these process types to determine the justificatory status of the belief? But it is irrelevant to my discussion how that question is answered. What is relevant is this: if the stories I mention are true, guessing is reliable and perception is unreliable. So if the stories are true then, on AWR, in *any* case where guessing is the relevant process, the output is justified; and in *any* case where perception is the relevant process, the output is unjustified. But it is false that in any case where these are the relevant processes, the outputs have those statuses. Yet we cannot simply deny the stories from the armchair. So we cannot accept AWR, on Enduring Reliability, Opportune Reliability, or Historical Reliability.

As foretold, AWR's problem worsens. While reflecting on only two belief-formation processes, guessing and perception, we found that several promising accounts of reliability poorly solve the temporality problem. Reflecting on other processes reinforces this result. As we identify paradigmatically unjustified beliefs formed by wishful thinking, hasty generalization, bias, etc., we can reason to conclusions similar to (C). As we identify paradigmatically justified beliefs formed by memory, testimony, introspection, etc., we can reason to conclusions similar to (C\*). But we wouldn't have reason to believe the consequent of these conclusions when AWR is paired with Enduring Reliability, Opportune Reliability, or Historical Reliability.<sup>16</sup> So we wouldn't have reason to believe the pairs.

Reliabilists may be tempted to insist, *à la* Kornblith (2004: 198), that justified belief just requires reliable belief-formation. It does not also require having access to any relevant external factors, or reason to accept the belief's consequences. So, we can reasonably accept one of the above pairs, even if we don't have reason to accept the consequent in (C) or (C\*).

Maybe something like this is generally correct. But in this context it harms more than it helps. Suppose justified belief in AWR just requires reliable formation. Anyone who justifiably believes AWR can still reasonably believe (P1) and (P1\*), and can reason along the lines above to (C) and (C\*), and then can *deduce* from AWR that the consequents of (C) and (C\*) are true. That is, anyone who has justified belief in AWR and has reasoned to (C) and (C\*) could simply *deduce* that no Blindluck-type scenario and no Experience Machine-type scenario occurs in the relevant temporal window. What's more, one could deduce this without having any idea what the relevant temporal window is! One could deduce that, whatever the window is, guessing's truth ratio in it is too low, and perception's is high enough. These deductions would justify. So, a strongly justified belief in AWR might allow one to *know*—even from the armchair, even with little information about what the world is, was, and will be like—that no Blindluck-type scenario and that no Experience Machine-type scenario occurs in the temporal window. It is absurd that we could acquire such empirical knowledge in this way. Enduring Reliability, Opportune Reliability, and Historical Reliability are inadequate solutions to the temporality problem, and the Kornblithian maneuver does not change that.

An issue-specific moral here: any *possible* reason to believe AWR will be reason to deny that a Blindluck-type scenario and Experience Machine-type scenario occurs during the relevant temporal window. So, the pool of potential reasons to believe AWR, when paired with one of the above accounts of reliability, looks dry. Maybe no such scenarios have occurred, and none ever will. But we receive little reason to believe none did or will *just* from the sort of support we seem to have for AWR—such as our limited observations about the relatively recent performance of

<sup>16</sup> Notably, most epistemologists of testimony agree with Lackey (2006: 440) that “the observational base of ordinary epistemic agents is simply far too small to allow the requisite induction about the reliability of testimony.” But our observational base for testimony is not relevantly smaller than our bases for perception, guessing, memory, hasty generalization, etc. If Lackey is right, then the consequents in conclusions like (C) and (C\*) lack support, and so AWR lacks support. If Lackey is wrong, then the current orthodoxy about testimony includes an importantly false assumption.

guessing and perception. A general moral here: when reliabilism allows for justified belief in the absence of relevant information, the use of that justified belief as a premise often allows for an unnatural deduction, that is, an allegedly justifying inference that yields an intuitively unjustified belief.

What I will call the *unnatural deduction problem* also arises for AWR when it's coupled with either of the remaining accounts of reliability. On Proximate Reliability, guessing is reliable at any time near a Blindluck-type scenario. AWR, on this account, implies that guessing justifies at those times. But, intuitively, guessing never justifies. So, if we have reason to believe AWR, we implausibly can deduce that no Blindluck-type scenario has obtained. Since we lack reason to believe the latter, we lack reason to believe AWR. Similarly, it is plausible that perception will never fail to justify. If that's right and AWR is true, then perception will always be reliable. So on Proximate Reliability, we can infer that no Experience Machine-type scenario will occur if we have reason to believe AWR. But we cannot learn about the future in this way, so we should not pair AWR with Proximate Reliability.

Period Reliability tells us that a process that performs adequately during a key period is reliable. Different versions of Period Reliability identify different periods. By now it is clear that any version identifying a period that includes the distant past or future introduces the unnatural deduction problem. Support for AWR, on those versions, lets us justifiably deduce the wrong kinds of empirical facts about those parts of the period. Versions of Period Reliability that identify a smaller, more local period are arbitrary, chauvinistic, and—if selected merely to avoid the unnatural deduction problem—ad hoc. The states of affairs that determine all justification happen to obtain in a fixed narrow period that we occupy and have the most information about. And the general problem still arises: anyone in the past and future with any reason to believe AWR could make unnatural deductions about the present. It seems that none of the accounts of reliability adequately solves the temporality problem.

One might think I have introduced no new problem for AWR because I merely raise a “problem of easy justification”.<sup>17</sup> This problem targets views that imply that experience by itself is a source of justified beliefs. It seems too easy on these views to justifiably believe that one is not being deceived. On these views, one can justifiably believe that *the table is red* if it merely appears so to one. But from this justified belief one could infer that *the table is not white with red lights shining on it* (i.e., that one is not being deceived). Although we can acquire justified belief in the latter, we can't clearly acquire it in this way. So appearances by themselves must not enable justified belief in the former.

I am not raising this problem. I am showing that AWR incorrectly allows us to justifiably believe certain empirical facts about the distant past and future, not about current deception. These are not clearly facts we already believe with justification. Also, my starting data are paradigmatic evaluations of beliefs, not appearances. And I only target views of justification which both allegedly accommodate this data and

---

<sup>17</sup> A variant of Cohen's (2002) problem of easy knowledge.

entail that justification depends on the empirical facts—i.e., a subset of externalist views. The problem of easy justification has a different target. The unnatural deduction problem introduces new challenges.

### 3.2 The difference-maker problem

The next problem arises for AWR on any account of reliability that meets a certain condition: on it, in some cases, the truth-value of a belief at a time helps shape whether its formation process is reliable then. That is, there are cases in which a process forms a belief at  $t$ , and  $t$  is *inside* the temporal window that determines the process's reliability at  $t$ . The five accounts of reliability from Sect. 2 meet this condition. On these accounts, AWR faces what I will call the *difference-maker problem*: a single belief ends up making differences in justification that it shouldn't. Here is one instance of this problem.

Whatever truth ratio we identify as minimally good enough for reliability, there will in fact be processes that barely meet it and others that barely do not. The truth-value of a single output of these processes in the relevant temporal window affects whether the process's truth ratio is good enough. It is plausible that there are processes bordering the threshold of reliability, because there are indefinitely many process types, with widely varying performances. Whatever the relevant temporal window is, the truth ratio of some process in it is good enough by a margin of just one true belief, and another is not good enough by a margin of just one false belief. For either process, on AWR, a single belief is a difference-maker for whether *all* beliefs formed by it in any world are justified (then).

The problem of interest here is not that AWR unintuitively attributes justificatory powers to certain processes and not others, or at certain times and not others. Rather, it is that AWR attributes powers for unintuitive reasons. It should not be that the truth-value of a *single* actual belief has made the difference in whether an entire process (namely, the one forming the belief) is justifying. Something so trivial as the truth-value of a single belief should not make a nontrivial difference in its formation process's truth ratio, such that all beliefs formed by that process share one justificatory status rather than another. It appears arbitrary that some processes are reliable in the actual world and are consequently justifying in all worlds. A trivial difference has made all the difference.

The difference-maker problem for AWR, on any of the five accounts of reliability, seems even more severe once we realize the following. When the truth-value of a single actual belief has made the difference in whether process  $r$  is justifying, then, if  $r$  is the process that settles that belief's justification, the truth-value of the belief makes the difference in whether it itself is justified. In such a case, the justificatory status of one's belief that  $p$  is explained by  $p$ 's truth-value, for *any*  $p$ . But justification should not depend so arbitrarily on truth. There may, of course, be some beliefs that are justified only if true. There is something special about the content of these beliefs. But when a belief's truth-value makes a difference to the reliability of its formation process, it does so regardless of its content. Unintuitively, the belief is justified because it happens to be true, or

unjustified because it happens to be false. The difference-maker problem renders AWR unacceptable on each account of reliability.

### 3.3 The instability problem

The final problem builds on the difference-maker problem, and applies to AWR when paired with a dynamic account of reliability. For simplicity, I will present the problem for AWR just on Historical Reliability.

On dynamic accounts, whether a process is reliable can change. The change coincides with changes in the process's truth ratio in the relevant temporal window—in particular, with changes in whether that truth ratio is good enough. Whatever the minimally good enough ratio is, there will in fact be processes hovering about it. This is plausible because, again, there are indefinitely many process types, varying broadly in performance. Whatever the relevant temporal window is, within it, there is a process with a truth ratio that at least briefly vacillates around what is minimally good enough. On the dynamic accounts of reliability, some process  $r$  is reliable at  $t_1$  but not at  $t_2$ , reliable at  $t_3$  but not at  $t_4$ , etc., because at the respective times  $r$  forms enough true beliefs, then enough false beliefs, and so on.

The *instability problem*, as I will call it, is that differences in justification over successive times could correspond to, and be explained by, unnoticeable minor fluctuations in the performance of a single process—fluctuations that result from just a few cases, far removed from the subject and her mental life. Earlier I supposed that dynamics in reliability trickle down to justification as follows: if process  $r$  forms a belief that  $p$  at  $t_1$ , the belief that  $p$  is justified at  $t_2$  just if  $r$  is reliable at  $t_2$  (a belief is justified just at the times its formation process is reliable). On this trickle-down theory, given the instability of process  $r$ 's reliability, a subject could form a belief that  $p$  via  $r$  at  $t_1$ , and the belief is justified at  $t_1$ , unjustified at  $t_2$ , and so on. The subject could be doing nothing relevantly different at these times. And as far as the subject can tell, there are no relevant changes at these times and her belief is equally reasonable at each. But on a dynamic account of reliability, AWR implausibly implies that every belief formed by  $r$  shifts justificatory status at those times, simply because of the truth-value of a few unheard of outputs at those times.

One might, in response, be tempted to adopt an alternative trickle-down theory that avoids this particular instability. A natural alternative is: if process  $r$  forms a belief that  $p$  at  $t_1$ , the belief that  $p$  is justified at any time just if  $r$  is reliable at  $t_1$ . That is, a belief's justification over time doesn't change, even if the reliability of the process that initially formed it changes. If dynamics trickle in this way, the belief formed by  $r$  at  $t_1$  can remain justified at  $t_2$ , even though  $r$  becomes unreliable at  $t_2$ .

Sadly, the instability problem still manifests itself on this trickle-down theory. Suppose that over  $t_1, t_2, t_3, t_4$ , etc.,  $r$  vacillates in reliability but forms beliefs that have identical or relevantly similar content. For example, Maria is staring at the setting sun, continuously forming the belief that *the sun is still setting*. Her belief formed at  $t_1$  is justified, her belief formed at  $t_2$  is unjustified, her belief formed at  $t_3$  is justified, her belief formed at  $t_4$  is unjustified, etc. But as far as she can tell she is using the same process to form (nearly) identical beliefs in the same circumstances,

and the beliefs are equally reasonable. But on a dynamic account of reliability, AWR implies that all beliefs  $r$  forms at  $t_1$ ,  $t_3$ , etc., differ in justification from those formed at  $t_2$ ,  $t_4$ , etc., because of the truth-value of outputs in some unheard of cases at those times.

This implication is implausible. And, if we deny this implication while retaining AWR and a dynamic account of reliability, we can deduce that no process's truth ratio has ever vacillated in the relevant temporal window! Perhaps there has never been such vacillation, but that's to be discovered empirically. We don't learn it from the armchair considerations that appear to support AWR and dynamic accounts.

In response, one might insist on absorbing the instability problem into the generality problem; when a process's reliability vacillates, its potential relevance for determining a belief's justification also wavers. More strongly: relevant processes are restricted to those that do not vacillate in reliability. A solution to the generality problem will also solve the instability problem.

Unfortunately, it is hard to see the appeal of this response. Part the motivation for pairing reliabilism with a dynamic account of reliability, presumably, is that doing so allows us to account for intuitions about how a process can have justificatory powers at one time but not another. But why select a dynamic account if mere changes in a process's reliability limit its relevance to settling a belief's justification? The motivation collapses. A process that can lose its justificatory power is, on this response, *irrelevant* to determining a belief's justification. Also, if relevance required unwavering reliability, then AWR would allow for inappropriate methods of empirical discovery. Because our intuitive, evaluative, higher-order beliefs (e.g., beliefs in (P1) and (P1\*)) seem justified independent of when formed, we could deduce that whatever process is relevant for their justification contingently does not oscillate in reliability and, therefore, that not too many of its recent outputs have been false. Finally, even if relevance wavers alongside reliability, I will still have shown that reliabilism is worse off than supposed. Solutions to the generality problem must meet even more demanding standards: they must be sensitive to mere temporary fluctuations in the reliability of potentially relevant processes. Since any commonsensical process could fluctuate in reliability, a solution to the generality problem must primarily involve processes that are much stranger than reliabilists have led us to believe. Reliabilism trades away much of its intuitive appeal.

I note that neither the difference-maker problem nor the instability problem disappears if we follow Goldman (2008a, 1979: 11) in claiming that the minimum truth ratio good enough for reliability is *vague*. If the minimum is vague, a single belief's truth value can decide whether the belief's formation process performs well enough to determinately justify. The very recent performance of the process can make its truth ratio cease to be determinately high enough, then be determinately high enough, etc., rendering vague whether the process justifies, and then rendering determinate whether it justifies, then vague, then determinate, etc. As an implausible result of unnoticeable minor fluctuations in a process's performance, there is instability in the determinate justificatory power of the process. What's more, if the minimum were vague, then on AWR whether perception and guessing have determinate justificatory powers would depend on whether these processes have performed well enough recently. And it is plausible that a given perceptual belief is



determinately justified, and that a given guess is determinately unjustified, regardless of when formed. We could therefore deduce that perception has performed well enough recently, that its less recent performance isn't bad, and that the same cannot be said for guessing. These conclusions may be correct, but this way of establishing them is illegitimate.

I have shown that none of the five most promising accounts of reliability helps AWR. There are no defended alternatives. The temporality problem is severe, and prospects for a solution are bleak.

#### 4 The generality of the reliability problem

The temporality problem is a reliability problem. It is the problem of providing principled and explanatory temporal parameters around the outputs that settle a belief-formation process's reliability at a time. We have found no parameters on which AWR seems true. I will show that no defended version of reliabilism fares better in the face of the reliability problem.

Each version generally falls into one of three classes, depending on how the version selects the non-temporal conditions that pick out the kind of tendency to produce true beliefs that matters for justification. I will briefly explain why problems arise on solutions for views within each class.

AWR belongs to a class we noted above:

*Lone World Reliabilism.* *S's belief that  $p$  is justified iff  $S$ 's belief that  $p$  is the output of a belief-formation process type that is reliable in  $w$ .*

One view in this class, which I will call *Same World Reliabilism*, identifies  $w$  as the world in which  $S$  forms the belief. A process's performance in a world determines its reliability there and thus, on Same World Reliabilism, determines the justificatory status of its outputs *there*. Besides AWR and Same World Reliabilism, no other kind of Lone World Reliabilism has been defended, and rightly so.<sup>18</sup> Of the indefinitely many other possible kinds, almost all seem arbitrary and implausible, since each identifies a justifying process as one that is reliable in some *single* non-actual world other than the subject's. Whatever world that is, it is hard to see why reliability in it could be so important. AWR and Same World Reliabilism are the only kinds of Lone World Reliabilism that might be right.

<sup>18</sup> Levin (1997) endorses Same World Reliabilism. For discussion see Lyons (2013). Sosa (2001) defends a view that entails both AWR and Same World Reliabilism but that assigns them distinct concepts of justification. Comesaña (2010, 2002) defends Indexical Reliabilism, a view intended to express both AWR and Same World Reliabilism; Ball and Blome-Tillman (2013) criticize it. On Indexical Reliabilism, the truth-value of all propositions expressed by a justification attribution depends on either of two worlds (the actual world and the subject's world), so it is not clearly a version of Lone World Reliabilism. At any rate, Indexical Reliabilism faces the main problems AWR and Same World Reliabilism face here since, on it, a process' actual performance determines both its reliability and its justificatory powers in the actual world (see below). Becker (2007: 32) attacks what he calls "actual world reliabilism" by citing an example of Plantinga's (1993: 199). Greco (2000: 175–176) also uses the example, but neither he nor Plantinga specifies whether it is meant to threaten Same World Reliabilism or AWR. The example involves a non-actual world, such that it threatens only Same World Reliabilism.

And Same World Reliabilism has no good solution to the temporality problem. While AWR and Same World Reliabilism differ, they share an important feature: the performance of a process in the *actual* world determines its reliability and thus its justificatory powers in the *actual* world. This feature was enough to prevent AWR from adequately solving the temporality problem. All defended versions of reliabilism with a feature like this are consigned to AWR's fate. A theory of justification should accommodate our data, including our intuitive evaluations of guesses, perceptual beliefs, etc. If Same World Reliabilism does this, it implies that no Blindluck-type scenario or Experience Machine-type scenario occurs in the relevant temporal window in our world. But we have already seen that we lack reason to believe this implication, on any account of reliability. Reason to believe Same World Reliabilism, on any account, would allow unnatural deductions about the actual past and future. Further, on Same World Reliabilism, implausibly, there is a process  $r$  that forms a belief  $b$ , the truth-value of which makes the difference in whether all  $r$ 's actual outputs, including  $b$ , are justified (at a time). And when Same World Reliabilism is paired with a dynamic account, the unstable reliability of some processes in our world results in troubling vacillations in justification in our world.

The next class of reliabilism schematizes most views that understand the tendency to produce true rather than false beliefs as a *propensity*, where a propensity has a modal measure:

*Modal Reliabilism.* S's belief that  $p$  is justified iff S's belief that  $p$  is the output of a belief-formation process that is reliable in  $W$ .

$W$  is a special, multi-membered domain of worlds. Modal Reliabilism is perhaps the most popular class, yet surprisingly few philosophers specify the special domain.<sup>19</sup> Goldman (1986) says it consists of normal worlds—worlds compatible with our general beliefs about our world. Henderson and Horgan (2010) say it consists of worlds that are experientially similar to, though perhaps otherwise different from, ours. William Alston (1995) and Goldman (2008b: 76n.7) are friendly to it consisting of the neighborhood of possible worlds near the actual world.

The conditions in which a process counts as reliable in a multi-membered domain are often left unclear. Goldman (1986: 395n.24) assumes that a process's truth ratio is "relatively invariant across" the worlds in his preferred domain, but this does not reveal the desired conditions. Maybe a process is reliable in a domain if it has a good enough truth ratio in *most* worlds in the domain. Or maybe if it has a good enough truth ratio in *each* world in the domain. Or maybe if the truth ratio of the process's pool of outputs from all worlds in the domain is good enough. Fortunately, any of these views is fine for my purposes.<sup>20</sup>

<sup>19</sup> Critics of Modal Reliabilism include Baumann (2009), Becker (2007), Frise (2014), Goldman (1988), Matheson (2015), and Vogel (2000).

<sup>20</sup> Henderson and Horgan (2010: 114–115) propose other conditions, centered on "degrees of variation" across the worlds in the domain, such that their version of Modal Reliabilism faces Probabilistic Reliabilism's reliability problem (see below).

If a Modal Reliabilist theory accommodates our intuitive evaluations about guesses, perceptual beliefs, and so on, then according to it, Blindluck-type scenarios and Experience Machine-type scenarios are sufficiently infrequent in the special domain of worlds. But the actual world is a normal world, and it is experientially similar to itself, and it is in the neighborhood of worlds near itself—it is in the domain. On all defended Modal Reliabilist views, features of the actual world partly determine whether processes used *in it* are reliable and justifying.

What's more, our general beliefs allow for a broad variety of pasts and futures, in which belief-formation processes perform with widely varying success; normal worlds are diverse. And, worlds experientially identical to ours have a broad variety of pasts and futures. And, our ignorance about the actual past and future extends to the pasts and futures of worlds neighboring ours. Within the special domain, either there is too much diversity or too much similarity to our world for us to suppose the domain contains sufficiently few Blindluck-type scenarios and Experience Machine-type scenarios. If we had reason to believe any Modal Reliabilist theory on any of the accounts of reliability, we could make unnatural deductions about not just our world, but also about other normal worlds, worlds experientially similar to ours, or worlds in the neighborhood—while nonetheless being fairly ignorant about which worlds these are!

In normal worlds, worlds experientially similar to ours, and worlds in the neighborhood of ours, some processes have a marginally high enough truth ratio. The truth-value of some single belief in one of those worlds is the difference-maker for whether all outputs of its formation process are justified (at a time) in our world. And some processes rapidly fluctuate in performance in these special domains, at least briefly. When a Modal Reliabilist theory is coupled with a dynamic account, the reliability of a process in the special domain can waver, such that justification has an implausible instability in our world.

A final class of reliabilism interprets reliability probabilistically:

*Probabilistic Reliabilism.* S's belief that  $p$  is justified *iff* the probability of  $p$ , given the process by which S forms belief that  $p$ , is sufficiently high.

Its defenders often understand all reliable, justifying belief-formation process types to be of the form *believing  $p$  based on evidence  $e$* , where  $e$  is the subject's evidence. It is crucial to clarify the intended sense of *probability*. If a version of Probabilistic Reliabilism interprets probability as something settled by a performance within a domain of worlds, then the version is ultimately a form of either Lone World Reliabilism or Modal Reliabilism, and so it lacks a solution to the temporality problem.

Alston (1988: 269) and Hill (1996: 577) intend the probability to be “objective,” but they elaborate no further. Comesaña (2010: 597n.37) is more informative, saying the probability relations are objective and “arguably necessary,” just as entailment relations are. The probability is logical, and not determined by any non-maximal region of modal space.<sup>21</sup> If S's belief that  $p$  is justified, it's because the

<sup>21</sup> Cf. Baumann (2009, 1998), Henderson and Horgan (2010), and Kvart (2006).

process of basing belief that  $p$  on  $S$ 's evidence  $e$  necessarily probabilifies  $p$ . Reliability and justification don't depend on the performance of a process in any temporal window in any worlds. Here, reliability might be best understood as a propensity, where a propensity is primitive, not to be reduced or conditionally analyzed.

For this reason, Comesaña's Logical Reliabilism (as I will call it) dodges the temporality problem. Still, Logical Reliabilism fails to solve the reliability problem. I claim that there is no *good* way of discerning the logical probabilities—discerning whether, for example, the probability of a guess is low, and the probability of a perceptual belief is high. So, it seems we cannot confirm Logical Reliabilism by testing its implications. And no other support for the theory seems available. Because the theory interprets reliability probabilistically, it is unsupported.

Here is why there is no useful guide to the probabilities. We might follow reliabilists like Alston (1995: 6) and use data about the performance of a process in some domain as a guide, even though reliability doesn't consist even in part in that performance. But then we run into a temporality-like problem. It is the problem of identifying the temporal parameters on the process's outputs that are a useful guide. A useful set of outputs will have many members, produced in a wide variety of circumstances. The variety helps correct for noise or misleading data. When, for example, we test whether induction probabilifies, we crucially rely on the performance of induction in a range of conditions. A better-than-chance performance in the range isn't what *makes* induction probabilifying, but it makes reasonable to believe that induction probabilifies. But what are the temporal parameters on the range of conditions that matters for discerning whether a process of belief-formation probabilifies? The need to answer this introduces riddles for Logical Reliabilism paralleling the riddles that the temporality problem introduced for the other forms of reliabilism.

A different guide might avoid that need: our intuitive evaluations of beliefs. If a belief seems justified, then we have reason to suppose it is produced by a process that makes it probable. But if such evaluations are the *only* guide, they are useless. We can no longer *confirm* Logical Reliabilism by testing its implications against our intuitive evaluations. We identified those implications simply by making them match our evaluations, so the match provides no evidence for Logical Reliabilism.

Logical Reliabilism attempts to solve the reliability problem by interpreting reliability probabilistically. Until reliabilists identify a useful guide to the relevant probabilities, Logical Reliabilism's attempt leaves it unsupported. The attempt introduces another problem: it forfeits a key externalist asset, such that (as I will show) the generality problem worsens. If probability relations are necessary and objective, they are invariant across circumstances. Reliability isn't indexed to worlds. Contingent environmental differences between possible subjects make no difference to whether the processes by which they form beliefs are reliable. How a process performs in an environment does not help explain why its outputs in *that* environment have their justificatory status. This renders irrelevant to epistemology much of the external world that externalists think matters.

How, then, is Logical Reliabilism even externalist—how could a contingent external condition make a justificatory difference? Perhaps the condition could still

somehow help select the *relevant* belief-formation process.<sup>22</sup> Contingent external features don't effect which processes are good, just which ones settle a belief's goodness. This surrenders more of reliabilism's merit to the generality problem. Reliabilism can no longer account for our intuitive evaluations of beliefs by citing environmental factors that affect how reliably the beliefs are formed. On the current proposal, reliabilism could only cite factors that affect which process is relevant in the circumstances. But if the reliability of a process is circumstantially invariant, it is hard to see why circumstances affect the process's relevance. It becomes more important, yet harder, to solve the generality problem well. Logical Reliabilism offers no good solution to the reliability problem. Maybe with development, a variant of Probabilistic Reliabilism will fare better. But I see little reason for optimism.

## 5 Conclusion

Externalists talk freely about reliability, but it is hard to understand that talk in a way that is friendly to their theories. The reliability problem plagues externalist views like process reliabilism that place reliability conditions on justification. And we have not found a promising solution. In its absence, we have new reason to resist reliabilism. Since only externalist views place reliability conditions on justification, non-externalist views do not face the reliability problem. We should count them as having an additional virtue, or as lacking yet another vice.

**Acknowledgements** I am grateful to Jon Matheson, Alyssa Ney, Brad Rettler, Lindsay Rettler, Brett Sherman, three anonymous referees, and an audience at UT Austin for valuable discussion of this paper. Thanks especially to Earl Conee, Rich Feldman, and Kevin McCain for helpful comments on multiple

<sup>22</sup> Comesaña (2006) attempts to solve the generality problem by identifying the relevant process as *basing belief that p on e*. But given his solution, contingent external factors cannot even affect which process is relevant. On his solution, Logical Reliabilism could remain externalist if these factors nonetheless determine which of a subject's mental features count as evidence for *p*. But if evidential support relations are necessary, then Comesaña's Logical Reliabilism paired with his solution to the generality problem would deny externalism. For further criticism of Comesaña's solution, see Matheson (2015).

The reader may have noticed that Comesaña (2010) endorses not only Logical Reliabilism but also Indexical Reliabilism, which expresses both Same World Reliabilism and AWR: roughly, *S's belief that p is justified iff the type producing a belief that p based on e is actually reliable*. Depending on how 'actually' is understood, this expresses either of two propositions—the horizontal or the diagonal proposition in 2-dimensional semantics. Comesaña says that the truth-values of the diagonal and horizontal propositions may differ. This is supposed to explain our allegedly conflicting intuitions about the deceived victim in the new evil demon problem; the victim's beliefs are justified (as AWR implies) because they are formed by processes that are reliable in the actual world, but they are unjustified (as Same World Reliabilism implies) because they are formed by processes that are not reliable in the victim's world.

Comesaña's reliabilism-cocktail has consequences he might not like. On Logical Reliabilism, the relevant probability relations are necessary—if a process is reliable in some world, it's reliable in all. So, the diagonal and horizontal propositions expressed by Indexical Reliabilism are logically equivalent. AWR and Same World Reliabilism, then, are identical. So the cocktail can't explain conflicting intuitions about the deceived victim.

drafts. I revised this paper while supported by a grant from the Templeton Religion Trust. The opinions expressed in this paper are my own and do not necessarily reflect the views of the Templeton Religion Trust.

## References

- Alston, W. (1988). An internalist externalism. *Synthese*, 74, 265–283.
- Alston, W. (1995). How to think about reliability. *Philosophical Topics*, 23(1), 1–29.
- Ball, B., & Blome-Tillman, M. (2013). Indexical reliabilism and the new evil Demon. *Erkenntnis*, 78(6), 1317–1336.
- Baumann, P. (1998). Can reliabilists believe in subjective probability? *The Philosophical Quarterly*, 48, 199–200.
- Baumann, P. (2009). Reliabilism—Modal, probabilistic, or contextualist. *Grazer Philosophische Studien*, 79, 77–89.
- Becker, K. (2007). *Epistemology modalized*. New York: Routledge.
- Bergmann, M. (2006). *Justification without awareness*. Oxford: Oxford University Press.
- BonJour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, 5, 53–73.
- BonJour, L. (2001/2010). *Epistemology: Classic problems and contemporary responses*. Lanham: Rowman & Littlefield.
- Cohen, S. (1984). Justification and truth. *Philosophical Studies*, 46, 279–295.
- Cohen, S. (2002). Bootstrapping and the problem of easy knowledge. *Philosophy and Phenomenological Research*, 65(2), 309–329.
- Comesaña, J. (2002). The diagonal and the Demon. *Philosophical Studies*, 110, 246–266.
- Comesaña, J. (2006). A well-founded solution to the generality problem. *Philosophical Studies*, 129(1), 27–47.
- Comesaña, J. (2010). Evidentialist reliabilism. *Noûs*, 44(4), 571–600.
- Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies*, 89, 1–29.
- Feldman, R. (1985). Reliability and justification. *The Monist*, 64(1), 59–74.
- Feldman, R., & Conee, E. (1985). Evidentialism. *Philosophical Studies*, 48, 15–34.
- Feldman, R., & Conee, E. (2002). Typing problems. *Philosophy and Phenomenological Research*, 65(1), 98–105.
- Frise, M. (2014). Speaking freely: On free will and the epistemology of testimony. *Synthese*, 191(7), 1587–1603.
- Goldman, A. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge*. Dordrecht: Reidel.
- Goldman, A. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.
- Goldman, A. (1988). Strong and weak justification. *Philosophical perspectives*, 2, 51–69.
- Goldman, A. (1992). Epistemic folkway and scientific methodology. In A. Goldman (Ed.), *Liasons: Philosophy meets the cognitive and social sciences* (pp. 155–175). Cambridge: MIT Press.
- Goldman, A. (2008a). Reliabilism. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Fall 2008 Edition). <http://plato.stanford.edu/archives/fall2008/entries/reliabilism/>.
- Goldman, A. (2008b). Immediate justification and process reliabilism. In Q. Smith (Ed.), *Epistemology: New essays*. Oxford: Oxford University Press.
- Goldman, A. (2009). Replies to discussants. *Grazer Philosophische Studien*, 79, 245–288.
- Greco, J. (1999). Agent reliabilism. *Philosophical perspectives*, 13, 273–296.
- Greco, J. (2000). *Putting skeptics in their place* (p. 2000). Cambridge: Cambridge University Press.
- Henderson, D., & Horgan, T. (2010). *The epistemological spectrum*. Oxford: Oxford University Press.
- Hill, C. (1996). Process reliabilism and cartesian scepticism. *Philosophy and Phenomenological Research*, 56(3), 567–581.
- Kornblith, H. (2004). Does reliabilism make knowledge merely conditional? *Philosophical Issues*, 14, 185–200.
- Kvart, I. (2006). A probabilist theory of knowledge. *Philosophy and Phenomenological Research*, 77(1), 1–43.
- Lackey, J. (2006). Knowing from testimony. *Philosophy Compass*, 1(5), 432–448.

- Levin, M. (1997). You can always count on reliabilism. *Philosophy and Phenomenological Research*, 57(3), 607–617.
- Lyons, J. (2013). Should reliabilists be worried about Demon worlds? *Philosophy and Phenomenological Research*, 86(1), 1–40.
- Matheson, J. (2015). Is there a well-founded solution to the generality problem? *Philosophical Studies*, 172, 459–468.
- Plantinga, A. (1993). *Warrant: The current debate*. Oxford: Oxford University Press.
- Pollock, J. (1984). Reliability and justified belief. *Canadian Journal of Philosophy*, 14(1), 103–114.
- Sosa, E. (2001). Goldman's reliabilism and virtue epistemology. *Philosophical Topics*, 29(1–2), 383–400.
- Sosa, E. (2003). Ontology, understanding, and the a priori. *Ratio*, 2, 178–188.
- Vogel, J. (2000). Reliabilism leveled. *Journal of Philosophy*, 97(11), 602–623.