

Does Reliabilism Have a Temporality Problem?

(~Penultimate Draft~, Forthcoming in *Philosophical Studies*, please *only* cite published version)

(word count: 6110)

1. Introduction

In a recent paper, Matthew Frise argues that reliabilist theories of justification have a *temporality problem* (2018).¹ He describes this as 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” (926). For example, if perceptual reliability determines whether a given perceptual belief B formed at t_1 is justified, one might wonder *which* perceptual belief-forming episodes fix the relevant truth ratio that determines B’s justification. Is it every case of perception from within 20 minutes of t_1 ? Every case of perception in the past? Every case of perception throughout all time? It’s unclear, initially, what to say here.

Frise considers a representative sample of principled temporal parameters that a reliabilist might invoke, and argues that there are serious problems with all of them. Frise constructs three distinct arguments to highlight the problems with these parameters: the unnatural deduction problem, the difference-maker problem, and the instability problem. He concludes that “the prospects for a solution [to the temporality problem] are bleak” (939). Moreover, Frise claims that the temporality problem is distinct from other problems with reliabilism, and that it constitutes a “*new* reason to avoid externalist theories that involve reliability in necessary or sufficient conditions for justification” (924, emphasis mine).

In what follows, I will examine Frise’s three arguments in turn. Ultimately, I argue that the temporality problem fails as a “new reason” to reject reliabilist theories of justification. As I’ll show, interpreting Frise’s arguments is a matter of some difficulty. I’ll argue that, for each of these three arguments, there are two importantly different interpretations. On one interpretation, the arguments against reliabilism invoke intuitively plausible premises. However, for these interpretations, I show that there is a reasonable temporal parameter that the reliabilist can adopt that allows her to successfully avoid the problem at the center of the argument. On the other interpretation of each argument, it’s less clear which temporal parameter might salvage reliabilism. However, this latter interpretation either invokes premises that are highly controversial and unmotivated, or invokes premises that simply re-raise substantive—and much older—disputes about reliabilism. In sum, Frise’s temporality

¹ Frise considers the temporality problem to be just one part of what he calls the *reliability problem*: “[T]he problem of supplying a suitable account of reliability” (924). Interestingly enough, a discussion of the temporality problem consumes almost his entire paper on the reliability problem. We can assume that he takes the temporality problem to constitute at least a significant part of the reliability problem.

problem faces a dilemma: either the reliabilist can straightforwardly defuse the problem, or Frise has failed to motivate any *new* reason to reject reliabilism.

2. The unnatural deduction problem

In order to demonstrate the intractability of the temporality problem, Frise presents several plausible representative temporal parameters that a reliabilist might consider. For our purposes, we need only consider two of them here:

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.

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. (926-8)²

These two parameters both specify a unique “window” of times that determine a process’s reliability (931). A process’s “window-truth ratio” is the reliability measurement that’s determined by its relevant temporal window.

Importantly, enduring reliability is a *static* parameter. According to static parameters, “a process is either reliable at all times or at none—its reliability does not change with time” (926).³ On the other hand, proximate reliability is *dynamic* in the sense that a process’s reliability can vary from time to time (928).

Frise points out that parameters can take both a modalized or non-modalized form. According to actual world reliabilism (AWR), a process’s degree of reliability is *rigidified* in the sense that it is determined by the process’s truth ratio across events occurring within the actual world (925). On the contrary, according to same-world reliabilism (SWR), the degree of reliability for a process occurring in world w is determined by its truth ratio *taken from that world w* (939). According to nearby-worlds reliabilism (NWR), the degree of reliability for a process occurring in world w is determined by a truth-ratio across belief-forming possibilities that are (counterfactually) “nearby” to w (940, 941).

Moreover, Frise explains how temporal parameters determine justification verdicts on particular cases only once they’re coupled with what he calls a “trickle-down” theory (928). Trickle-down theories explain how the reliability of a process determines whether a given belief is justified at a given time (928). Frise presents two candidate trickle-down theories, which I formulate as follows:

Exactly While (EW)	If process r produces a belief that p at t_1 , the belief that p is justified at some other time t_2 just if r is reliable at t_2 .
--------------------	---

² See (926-8) for Frise’s presentation of three other parameters.

³ Enduring reliability is static because the truth ratio mentioned in the consequent must be high *for any* time t , not just some finite range of times or some particular time.

Time of Formation (TF) If process r produces a belief that p at t_1 , the belief that p is justified at any time just if r is reliable at t_1 . (928)⁴

Frise first shows, in depth, how various parameters, once coupled with AWR and EW, are all undermined by the three arguments that highlight the temporality problem.⁵ Then, he argues that similar problems arise once these parameters are combined with other approaches like SWR and NWR. I'll begin by discussing Frise's presentation of the unnatural deduction problem for reliabilism.

We can briefly state the unnatural deduction problem for reliabilism as follows. First, consider plausible claims about justification like, "Typical beliefs based on perceptual experience seem justified. Guesses don't" (923). Frise contends that the conjunction of plausible claims like these *and* reliabilism—as specified by any candidate parameter and trickle-down theory—entails empirical claims about the distant past or future. As a result, insofar as we're justified in believing reliabilism as specified by any of these parameters and trickle-down theories, we would be able to justifiedly deduce empirical claims about the distant past and future simply on the basis of these propositions about justification and reliability (933). But Frise points out that this is an "absurd" method for making judgments about the distant past or future—a method that could never yield knowledge or justification (934). In other words, reliabilism licenses unnatural deductions of this sort. To avoid licensing such deductions, Frise suggests that we reject reliabilism.

For simplicity, I'll just present Frise's formulation of the unnatural deduction problem starting from the premise that perceptual beliefs are justified.⁶ Where "B" represents any perceptual belief—i.e., a belief that "takes perception at face value"—Frise claims, as a first premise, that we have good reason to believe this proposition:⁷

P1* B is justified (931).

Frise notes that "In order for AWR to be correct, it must accommodate our having justified beliefs in things like...(P1*)" (931). Frise then claims that,

⁴ As an anonymous referee insightfully points out, EW and TF are rather odd in how, given their current formulation, they're principles describing how a given belief b might have justification *at any time* – including times at which the belief b hasn't even been formed yet. The referee, correctly, points out that these principles misrepresent the notion that reliabilists (and other epistemologists, for that matter) actually care about analyzing: whether a given belief b has justification *at times when b exists/is being held*. As the referee notes, Frise could accommodate this concern by adding the further necessary condition in EW and TF to specify that "the belief that p exists" at the given time in question. Furthermore, it does not appear that this modification would affect the cogency of Frise's arguments in any way.

⁵ As I'll note later, Frise does not offer a lengthy discussion of reliabilism's ability to handle the unnatural deduction problem when combined with the TF trickle-down theory. This is because he thinks that "[t]here 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 [EW], since the other [trickle-down theories] in the end aid [reliabilism] no better" (928).

⁶ In his essay, Frise also presents a version of the argument that begins with the premise, *Guesses are unjustified*. This version of the argument proceeds in a structurally analogous manner, so I've omitted this version for brevity's sake.

⁷ Frise identifies the process type *perception* with "taking perception at face value" (931).

By understanding AWR, we can reasonably believe:

P2*. If AWR is true and B is justified, then perception is reliable. (931)

Next, Frise points out that the following proposition straightforwardly follows from reliabilism:

P3*. If perception is reliable, then perception's window-truth ratio is good enough. (931)

At this point, Frise argues that—from these three claims—one can thereby deduce empirical claims about distant points in time. He asks us to consider a hypothetical scenario.

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. (931-32)⁸

If any scenario like *Experience Machine* obtains within the temporal window for B, then its window-truth ratio will be very low. Frise points out that we can then infer the following claim by reflecting on the nature of reliability:

P4*. If perception's window-truth ratio is good enough, then no Experience Machine-type scenario occurs in the window. (932)

Frise notes that P2*, P3*, and P4* entail,

C*. If AWR is true, then no Experience Machine-type scenario occurs in the window. (932)

But now, according to Frise, it's clear that AWR—once coupled with P1*, various parameters, and various trickle-down theories—allows us to deduce surprising empirical claims. First, consider the enduring reliability parameter. Enduring reliability invokes a truth ratio taken across every moment of time—past, present, *and future*. But if this is the case, then P1*-P4*, C*, enduring reliability, and AWR entail that an experience machine-type scenario will *never*

⁸ For the version of the argument that proceeds from the premise, *guesses are unjustified*, Frise presents a hypothetical scenario called *Blindluck*. In this scenario, it turns out that for a large part of the history of the evolution of the human cognitive system, "guesses," (i.e., beliefs formed without the basis of consciously accessible information and phenomenology) constituted the majority of beliefs being formed, and that they were for the most part correct (because they allowed us to survive as a species). Hence, for a very large temporal window, *guessing* had a high truth ratio (931-2).

occur in the future (933).⁹ After all, if it does occur in the future, then the reliability of perception takes a serious hit according to the enduring reliability parameter.

Putting it all together, Frise notes that justifiedly believing the conjunction of P1*, AWR, and enduring reliability, would allow one to justifiedly deduce—“even from the armchair”—that an experience machine-type scenario *will never* occur (934). As Frise points out, this is a preposterous method for forming such a belief about the distant future—a method that could never yield justification (934). To avoid licensing this unnatural deduction, Frise suggests that we reject the conjunction of AWR and enduring reliability.

According to Frise, a similar problem plagues other parameters like proximate reliability. Frise writes,

[I]t 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. (935, emphasis mine)¹⁰

3. A response to the unnatural deduction problem

It’s important to reiterate that the unnatural deduction problem plagues specific and complex *conjunctions* of theses. These conjunctions include a statement of a parameter, a trickle-down theory, and a statement of a modal (or non-modal) reference class, etc.¹¹ Of course, Frise does not discuss *every* possible combination of theses that might constitute a reliabilist theory of justification. For parameters, he presents a representative sample of five that he takes to be “at least as plausible as any” (926). After showing how some conjunctions of views employing these five parameters fall prey to the variants of the temporality problem, Frise concludes that “we have not found a promising solution [to the temporality problem]. In its absence, we have new reason to resist reliabilism” (943).

⁹ Given the nature of enduring reliability, this conclusion follows regardless of whether EW or TF is the operative trick-down theory.

¹⁰ One might think that the unnatural deduction problem for AWR merely constitutes a good reason for the reliabilist to opt for a *modalized* interpretation of reliabilism’s temporal parameter, like NWR. However, Frise argues that the unnatural deduction problem plagues these interpretations of reliabilism as well.

On all defended Modal Reliabilist views, features of the actual world partly determine whether processes used in it are reliable and justifying... 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... 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 deduction about not just our world, but also about other... worlds in the neighborhood—while nonetheless being fairly ignorant about which worlds these are! (941)

So, according to Frise, the unnatural deduction problem is actually worse for NWR. Not only could we implausibly come to know obscure facts about our world, but we can also gain a sort of “insight” into innumerable neighboring worlds.

¹¹ As I discuss in section 7, this conjunction also includes a theory of *process type relevance*.

However, I think there is a conjunction of theses available to the reliabilist that successfully avoids the unnatural deduction problem. This conjunction of views centrally includes a commitment to proximate reliability and the TF trickle-down theory. But before showing how this conjunction avoids licensing unnatural deductions, I must first address a crucial move in Frise's presentation of the unnatural deduction problem for proximate reliability.

Frise contends that, "intuitively, guessing *never* justifies," and that "[s]imilarly, it is plausible that perception *will never fail to justify*" (935, emphasis mine). For simplicity, I'll frame this premise in the unnatural deduction for proximate reliability as follows:

PJ Perception will never fail to justify.

First, it's not immediately clear what Frise means by PJ. We can highlight at least two importantly different readings of this sentence:

PJa There will never be a time at which *presently* (or near presently) held beliefs formed by perception fail to be justified.

PJb There will never be a time at which any perceptually-formed belief—formed and held at *any time*—fails to be justified.¹²

While PJa is stronger than P1* in how it specifies that current perceptual beliefs will not lose their justification *at future times*, I'll concede that it remains roughly as plausible as P1*. However, PJa fails to generate an unnatural deduction problem for reliabilism once combined with TF and proximate reliability.

The selection of a trickle-down theory is essential for determining the implications of PJa. If we adopt proximate reliability and EW, committing to PJa still leads to an unnatural deduction about the distant future. According to EW, whether a given belief *b* formed at time t_1 by process *r* is justified at any other time t_2 depends on whether the process *r* is reliable as measured *relative to the time* t_2 . If proximate reliability is the correct parameter, then, according to EW, a process *r*'s unreliability around some future time t_2 can make it the case that beliefs formed by *r* at some earlier time t_1 lose their justification *at* t_2 . Suppose that an experience machine scenario occurs 70 years in the future. On EW and proximate reliability, the perceptual beliefs of today's 15-year-olds, that they are forming *now* and storing in memory, will lose their justification once they turn 85 simply because the process used to form those beliefs 70 years prior has become unreliable given the (then) current state of the world. These stored beliefs will lose their justification even if they've been properly stored in memory for 70 years and even if the subjects haven't acquired any defeaters for them throughout the 70 years. It's unreasonable that one's perceptual beliefs *formed today* could eventually lose their justification in this fashion. However, insofar as one is committed to PJa, EW, and proximate

¹² For clarity's sake, I'm building in to these two readings that the agent stores her perceptually formed beliefs in memory in a justification-enabling fashion, and that the agent never acquires defeaters for her perceptually formed beliefs held in storage.

reliability, one can thereby—unnaturally—deduce that no experience machine scenario will obtain in the distant future.

But proximate reliability and PJa, once coupled with TF, does *not* lead to an unnatural deduction about the distant future. According to the conjunction of TF and proximate reliability, it's only perception's truth-ratio from around time t_1 that determines whether a perceptual belief formed at t_1 has justification *at any time*. Thus, even if an experience machine scenario obtains 70 years from now, the current perceptual beliefs of today's 15-year-olds *won't automatically* lose their justification once they turn 85 simply because most uses of perception will (at that time) be hallucinatory—which is intuitively the right result. PJa, proximate reliability, and TF only entail empirical claims (pertaining to the reliability of perception) regarding the period of time *close to the present*.

Here, we see a critical error in Frise's presentation of the unnatural deduction problem. He only discusses the implications of proximate reliability when combined with the EW trickle-down theory because he thinks that “the other [trickle-down theories] in the end aid [reliabilism] no better” (928). But I have just shown this to be false. Proximate reliability and PJa, *once combined with TF*, don't entail any interesting empirical claims about the distant past or future.¹³

By Frise's own lights, this is serious blow to his argument. At the end of his presentation of the unnatural deduction problem, Frise re-affirms that “I am showing that [reliabilism] incorrectly allows us to justifiedly believe certain empirical facts about the distant past and future, *not about current deception*” (935, emphasis mine). In other words, Frise thinks that reliabilism's licensing unnatural deductions about distant points in time is *central* to the strength of his argument against reliabilism. If all he could show is that reliabilism (once filled in with a plausible parameter and trickle-down theory) allows one to deduce empirical claims about the present, this would make for a much less compelling reason to reject reliabilism.

But one might think that reliabilism's licensing of deductions pertaining to present empirical matters is still problematic.¹⁴ Consider the claim E: *A widespread experience machine-type scenario does not obtain during the period of time close to the present*. Isn't it unreasonable to think that we could justifiedly learn empirical truths like E “from the armchair” simply by reflecting on commonsense claims about justification and epistemological theses like PJa, TF, AWR, and proximate reliability?

¹³ The reliabilist can defend proximate reliability from the version of the unnatural deduction argument that proceeds from the premise *guessing is unjustified* or *guessing never justifies* in an analogous fashion. However, there might be an even simpler way to show that *guessing never justifies* fails to lead to an unnatural deduction.

Frise describes a guessed belief as one that “seems to the subject formed arbitrarily, seems formed without enough relevant information, without even a hunch or a memory of a relevant good track record” (930-1). As Frise notes, some reliabilists view reliability as merely a *necessary* condition on justification. Some epistemologists hold that a belief must be based on the right sorts of phenomenological or doxastic *grounds* in order to have justification. For example, Robert Audi claims that the only justified non-inferential beliefs involve some sort of intuitive, perceptual, introspective, or memorial phenomenology (2011: 6-7). Guesses, according to Frise's description, lack these sorts of grounds. Hence, Reliabilists who hold to the *right grounds* necessary condition as well can account for the claim *guesses never justify* without even invoking the unreliability of any sort of process—thus blocking an unnatural deduction.

¹⁴ Much thanks to an anonymous referee for raising this reasonable defense of the unnatural deduction problem.

In response, I contend that reliabilists are in a position to agree that *purely armchair* reasoning to empirical claims like E does not yield justification. This is because E, in the envisaged example above, is deduced from *the conjunction* of PJa, AWR, proximate reliability, and TF. Presumably, deductive inferences yield doxastic justification only if the subject justifiably believes each of the premises comprising the inference.¹⁵ Here, reliabilists will deny that we justifiably believe general claims like PJa simply by reasoning from the armchair. From a reliabilist perspective—assuming AWR—one could justifiably believe PJa only if she formed this belief by using a reliable process that was in some way sensitive to the present truth-conductivity of perception across the (actual) world. But it’s implausible to think that humans possess an “armchair-reasoning” cognitive mechanism that is sensitive to the global performance of perception in this way. Reliabilists will claim that, in order to reliably form a belief in PJa, humans need to gather and invoke at least some empirical evidence about one’s external environment and the workings of perception across the world.¹⁶ Hence, reliabilists have the theoretical resources to deny that proximate reliability and TF license deductions to empirical claims like E *from purely armchair reasoning*.

Given Frise’s own perspective on the unnatural deduction problem, it makes most sense to interpret his argument against the proximate reliability parameter as follows: a justified belief in proximate reliability, combined with any candidate trickle-down theory, would allow one to make justified deductions about *distant* points in time given that one also justifiably believes PJ. As I’ve shown above, in order for this conclusion to follow, we must interpret PJ as PJb rather than PJa. PJb says that any use of perception *from any period of time*—not just the present—will always justify. This being the case, if we are somehow antecedently justified in believing PJb, then invoking proximate reliability and TF *would* allow us to deduce that there will be no future experience machine-type scenarios. Hence, invoking proximate reliability and TF will still license unnatural deductions about distant points in time for one who justifiably believes PJb.

But PJb is much stronger than both P1* and PJa. Frise doesn’t offer any reasons in favor of PJb, let alone any discussion of how PJ ought to be interpreted.¹⁷ But given that PJb is a natural interpretation of PJ, as well as the fact that PJb delivers Frise’s desired conclusion, it’s worth considering what one *might* say in defense of PJb. Importantly, various contemporary conceptions of justification could be marshalled in support of PJb. First, one might think that a given belief is justified if and only if it’s formed in a way that doesn’t violate any of the subject’s epistemic duties. Call this the deontological conception of justification.¹⁸ One might think that, *ceteris paribus*, there will never be a time at which taking perception at face value to form simple perceptual beliefs fails to satisfy one’s epistemic duties. Were this the case, then PJb—or something close to it—would be correct. Secondly, one might think that a given belief is justified if and only if its content is the best explanation of the evidence on which the

¹⁵ See Goldman (1979) for a reliabilist-friendly formulation of this principle of inferential doxastic justification.

¹⁶ This could include basing a belief in PJa on expert testimony. Presumably, testimonial belief-formation is not a form of armchair reasoning.

¹⁷ Frise claims that it is “intuitive” that both guessing never justifies and that perception will always justify (935). But the PJb rendering of this sentence is an *eternal* truth claim—one that characterizes all times in the future—and it’s not clear why we should think that we could “intuit” a claim like this.

¹⁸ See Alston (1985: 60-8) and (1988) for an in-depth discussion of the deontological conception of justification.

belief is based. Call this the best-explanationist conception of justification.¹⁹ Furthermore, one might think that the simple contents of perceptual experience will always, *ceteris paribus*, be the best explanations of the perceptual experiences themselves. These claims would support P**J**b as well, given that believing the contents presented to oneself in perceptual experience is how one “takes perception at face value.”

For our purposes, it bears pointing out that invoking either of these two conceptions of justification in favor of P**J**b would make for a very controversial rendering of the unnatural deduction problem. Both deontological and best-explanationist conceptions of justification have faced thorough criticism.²⁰ Moreover, reliabilists think they have independent reason to accept an alternative conception of justification: a belief is justified if and only if it is formed in a *truth-conducive* manner.²¹ In addition, reliabilists have arguments for identifying process reliability as the relevant sort of truth-conductivity that generates justification.²²

As things stand, it’s unclear why reliabilists should believe something as strong as P**J**b. This being the case, the unnatural deduction problem is insufficiently motivated to the extent that it relies on P**J**b as a premise.²³ On the other hand, P**J**a is much more reasonable. But as I’ve argued above, if P**J**a is the correct rendering of the P**J** premise in Frise’s argument against proximate reliability, then the reliabilist who accepts TF has a ready response that blocks unnatural deductions about the distant past or future. Either way, we currently lack sufficient reason to believe that every candidate parameter licenses unnatural deductions about distant points in time.

¹⁹ See Conee and Feldman (2008: 97-99) and McCain (2014: 117-156) for a defense and presentation of a best-explanationist conception of justification.

²⁰ For example, Alston (1988) argues that the deontological conception is inconsistent with plausible principles of doxastic involuntarism. Goldman (2012: 147-149) argues that there are many kinds of justified belief (e.g., introspective justification, memorial justification, arithmetical inference-based justification) for which best-explanationism cannot account.

²¹ As opposed to the deontological conception, Alston argues that we should prefer a conception according to which being “justified in believing that *p* is to believe that *p* in such a way as to be in a strong position thereby to attain the truth and avoid error. It is to believe that *p* in a ‘truth conducive’ way” (1986: 193). Similar to Alston, as opposed to best explanationism, Goldman suggests that we adopt a “veritistic” conception of justification (2012: 147). According to veritism, the fundamental, or, *cardinal* epistemic value is *true* belief (1999).

²² For example, see Goldman’s classic arguments in his (1979). Elsewhere, Goldman argues that “The obvious strategy for veritistic unitarianism is to defend a reliabilist theory of justification, or at least some form of truth-linked justification theory...Beliefs are regarded as justified when they are produced by these very truth-conducive processes” (2002: 53-4).

²³ Reliabilists will be apt to point out that, *insofar as* we have good reason to believe reliabilism, we also have good reason to believe that justification for P**J**b is very hard to come by. According to reliabilism, process types like perception generate justified belief *only so long as* they are sufficiently reliable. In order to justifiedly believe something as strong as P**J**b, the reliabilist will demand that we use some reliable belief-forming process that is sensitive to the truth-conductivity of perception throughout *all* future points in time. But it’s implausible that we have ready access to any sort of belief-forming process like this.

4. The difference maker problem

The second argument aimed at demonstrating the temporality problem is what Frise calls the difference maker problem. He claims that this problem arises on any account of reliability according to which, “in some cases, the truth-value of a belief at a time (t) helps shape whether its formation process is reliable (at t)” (936). Frise asks us to imagine “[processes] bordering the threshold of reliability” such that one more true belief within the temporal window would suffice for pushing the reliability of the process just past the threshold for justification (936). But then, according to Frise, *the truth* of a belief arising from such a process plays a role in explaining the justification of that belief—a role that, according to Frise, it shouldn’t play:

Something so trivial as the truth-value of a single belief should not make a nontrivial difference in its formation process’s truth ratio...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...Unintuitively, the belief is justified *because it happens to be true*, or unjustified *because it happens to be false* (936-7, emphasis mine)

Importantly, Frise later considers the possibility that the reliability threshold for justification could be vague (938). But he claims that it’s just as implausible to countenance belief-forming processes that go from determinately unreliable to indeterminately reliable on account of one true belief, or processes that go from indeterminately reliable to determinately reliable on account of one true belief.

5. A response to the difference maker problem

At first glance, the main thrust of Frise’s objection seems to be that “trivial features”—like a mere true belief—cannot make “non-trivial” differences pertaining to the justification of the belief. Before proceeding, it bears pointing out that virtually *any* theory that takes justification to have a *threshold* of some sort could fall under the charge of having trivial features making non-trivial differences. Consider a classic case of induction. Presumably, after S has looked at one green emerald, it doesn’t seem as if S is justified in believing that q: *All emeralds are green*. The same goes for looking at 2 emeralds, 3 emeralds, etc. But, reasonably, there comes a point at which, after observing N green emeralds, S’s evidential base E_g *determinately* justifies her in believing q. But then this means that a seemingly trivial feature—the difference between observing N emeralds and $N-1$ emeralds—makes a non-trivial difference to the justification of q. Presumably, any reasonable theory of justification—evidentialist, reliabilist, or otherwise—should account for this phenomenon.²⁴ We could multiply cases like this, which suggests that

²⁴ Consider Conee and Feldman’s account of evidentialism, according to which believing p is justified for S iff believing p fits S’s evidence, where some p fits a body of evidence E iff p is the best explanation of E (2008:97-

we should charitably interpret Frise as *not* rejecting the general claim that justification has a threshold nature. Rather, it's important to focus on Frise's qualms with the explanatory role *truth* might play in justification.

At first glance, it's unclear which principle Frise takes to motivate the difference maker problem, as there are at least two reasonable interpretations of this principle from Frise's text above.

- D1 For any belief B, the truth of B cannot possibly constitute either the *only* explanation or the *main* explanation for B's justification.

- D2 For any belief B, the truth of B cannot possibly constitute *any part* of the explanation for B's justification.²⁵

Adopting the D1 interpretation of Frise gains its plausibility from his *unqualified* use of "explained by" and "because" in the above quotation.

I'm prepared to concede that D1 is quite reasonable on the grounds that, upon reflection, we see that there is a significant difference between a belief's being true and a belief's being justified. It's easy to think of true beliefs one might have that fall *far* short of being justified. The ease of multiplying cases like this at the very least provides strong *prima facie* support for a general principle like D1.

However, D1 is of no use—in this context—for an argument against reliabilism. The reliabilist is by no means committed to the thesis that the truth of a given belief can constitute the main explanation for its justification. According to reliabilism, the truth of the target belief will only be a very small *part* of the explanation of its justification. The reliability of the belief-forming process will depend on the truth of the process's outputs from *all* of its instances within the relevant proximate window. This is not importantly different than viewing the Nth observance of a green emerald as one small part (along with all the other observances) of the explanation for q's justification.

Before considering D2, it's important to note, along with Frise, that some forms of reliabilism are consistent with D2. Take NWR, according to which a process's track record across a reference class of *nearby possibilities* determines justification. Interestingly, Frise suggests that one can easily extend the difference-maker problem to NWR, seeing as how in "[possible] worlds in the neighborhood of ours, some processes have a marginally high enough

8). Reasonably, there comes a point at which, after observing N number of green emeralds, S's evidential base E_g is *determinately* best explained by q (*all emeralds are green*). But this means that the trivial feature of observing one green emerald (the difference between N and $N-1$) makes a non-trivial difference pertaining to the justification of q for S according to Conee and Feldman's evidentialism.

²⁵ For brevity, I've left off a relevant qualification of Frise's claim. More precisely, he grants that there can be some cases in which the truth of a belief constitutes a *content-sensitive* explanation for its justification (936). For example, consider an occurrent demonstrative belief with content that (successfully) refers to an occurrent experience, like, *I am having this experience*. A belief with this content might very well be justified (at least partially) in virtue of the successful reference of the thought, and hence, in virtue of the truth of its content. Importantly, Frise takes reliabilism to be committed to the thesis that the truth of a belief can constitute a *content-neutral* explanation of its justification, and it's this commitment that's, allegedly, so counter intuitive. For ease of presentation, I've left the "content neutral" qualifier out of D1, D2, and D3. However, following Frise, I am assuming a content-neutral notion of explanation in my formulations of these principles.

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” (941). Frise seems to find this modal sort of justificatory “difference-making” just as problematic as the original sort. Hence, we could interpret Frise as invoking the following principle against NWR.

- D3 For any belief B, the truth of a nearby possible belief relevantly similar to B cannot possibly constitute *any part* of the explanation for B’s justification.

However, similar to what we saw with PjB, D2 and D3 are substantive theses and ill-suited to serve as *starting points* in an argument against reliabilism. Affirming D2 and D3 is tantamount to denying reliabilism’s central and unique contribution to a theory of justification, which we can formulate as follows:

- TR Epistemic justification is at least partially grounded in a *truth ratio* fixed by a reference class of events (possible or actual) that bear important *causal* similarities to the token belief-forming process.

Many epistemologists—not just reliabilists—want their theory of justification to provide *some* informative explanation of how justification is connected to truth.²⁶ Reliabilists see TR as articulating the truth-connection in an elegant and explanatorily powerful way. Frise presents no defense of D2 and D3, so the difference maker problem will have little pull on those antecedently attracted to reliabilism. Furthermore, D2 and D3 don’t appear to be self-evident or axiomatic principles of justification. They seem no more reasonable than their negations at first glance, so any version of the difference maker problem that takes them as *premises* won’t convince those antecedently agnostic about reliabilism as well. Most importantly, simply denying the unique reliabilist approach to the justificatory truth-connection merely *re-raises* an older debate about the nature of this truth connection.²⁷ It doesn’t constitute a *new* reason to reject reliabilism.²⁸

²⁶ As Alvin Goldman notes, many theories of justification—both internalist and externalist—are keen to posit *some* sort of connection between truth and justification (2002). However, reliabilism uniquely analyzes this truth connection in terms of the kinds of truth ratios mentioned in TR.

²⁷ Contrast the TR account of the truth connection with Earl Conee’s evidentialist account: “Epistemic justification of a proposition is evidence of its truth. The relation of evidential support is the truth connection” (1992: 667). Conee clarifies that the relation of evidential support “is not obviously within the ontology of any current or prospective science” (668). One might have thought that *evidential support* could be analyzed in terms of naturalistic concepts, but Conee admits that “no such reduction seems to be in the offing for the relation of giving evidence” (668). This is in sharp contrast to how most reliabilists view the truth ratios mentioned in TR—namely as invoking naturalist-friendly notions like *event* and *causation*. For further discussion of the relationship between naturalism and reliabilism, see Goldman (1994).

²⁸ Frise also hints at a related but different sort of difference maker problem that specifically plagues AWR: 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. (936, emphasis mine)

The implication articulated in the italicized portion only follows on AWR, which rigidifies the reliability and justification-conferring properties of process types. This implication does not obtain on SWR and NWR.

6. The instability problem

The third argument that raises the temporality problem is what Frise calls the instability problem. The instability problem only plagues dynamic temporal parameters like proximate reliability. Hence, addressing the instability problem is relevant to my proposed solution to the unnatural deduction problem. According to Frise,

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* (937, emphasis mine).²⁹

Frise asks us to imagine a possible scenario in which Maria is watching the sun set slowly over the horizon. Using perception, she forms the distinct yet similar beliefs that *the sun is still setting* at time t_1 , t_2 , t_3 , t_4 , etc. (937). However, unbeknownst to her, the process type *perception* undergoes significant changes in its truth-ratio corresponding to its relevant temporal windows from t_1 to t_2 , from t_2 to t_3 , etc. As a result, her *the sun is still setting* belief is justified at t_1 , but then unjustified at t_2 when she forms it. Later on, it's justified at t_3 , but then unjustified at t_4 , etc. Frise notes,

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 (937, emphasis mine).

Because dynamic parameters like proximate reliability allow the possibility of such changes in justificatory status, we should reject them.

7. A response to the instability problem

From the above quotations, there appear to be two distinct sources of conflict for proximate reliability: the *remoteness* of factors contributing to justificatory status and the reflective *inaccessibility* of factors contributing to justificatory status. I'll consider both sources in turn.

First, Frise denies that events “far removed” from the subject and actual belief-forming process can have any sort of impact on the belief's justificatory status. If we interpret “far removed” both spatially and causally, I think we can conceive of cases that support Frise's

Hence, insofar as one finds the italicized implication too problematic, one can take this consideration as merely a reason to reject AWR rather than reliabilism per se.

²⁹ Frise presents versions of this problem for proximate reliability combined with either EW or TF (937). I'll simply discuss the latter formulation, seeing as how proximate reliability combined with TF was shown to present a plausible solution to the unnatural deduction problem in section 3.

concerns. Suppose that, right now, unbeknownst to us, in a far-off galaxy, there is an enormous population (in the trillions) of agents like us who form perceptual beliefs. However, almost all of them are brains in vats forming mostly *false* perceptual judgments. As a result, the process type *perception* has a very low truth ratio with respect to *all actual* cases of perception occurring within a proximate window of time, even though the truth ratio of perception taken across proximate cases *on Earth* is very high. Intuitively, the unreliability of perception, conceived of *universally*, doesn't seem to undermine the justification of everyday perceptual beliefs currently occurring on Earth.

I think that there are two principled ways in which reliabilism can be precisified so as to accommodate the intuitive verdict on this sort of case. The first solution involves a partial response to the famed generality problem. The generality problem asks, "Out of the innumerable process types exemplified by a given process token, *which* type is relevant for determining whether a belief is formed with justification-enabling reliability?"³⁰ After all, it seems that general repeatable types are the entities measured for reliability, not the one-off tokens. Frise notes that the generality problem, technically speaking, is distinct from the temporality problem, and that answers to one don't necessarily constitute answers to the other (929). However, I do think that certain answers to the generality problem can alleviate the instability problem for proximate reliability.

For example, the reliabilist could claim that relevant types are *community-relative*. On this account, whether subject S's perceptual judgments have justification depends on the truth ratio of all perceptual judgments throughout S's community within the temporal window.³¹ According to this (partial) answer to the generality problem, the track record of the spatially far-off envatted brains would be irrelevant to the justification of perceptual judgments made by members of communities here on Earth.

The second solution is to simply adopt a modalized conception of reliabilism like NWR.³² Clearly, possible belief-forming events performed by envatted individuals in far off galaxies are *counterfactually* very far away from perceptual belief-forming events currently taking place here on earth. According to NWR, these counterfactually far-off possibilities will be irrelevant to the justification determining truth-ratios of perceptual beliefs taking place here on earth.

If solving the instability problem is merely a matter of ruling out "far away" factors from determining justification, then then these two solutions show that the reliabilist has conceptual resources for defusing this objection in a non-ad hoc manner. However, given Frise's comments pertaining to what "the subject can tell" and factors "far removed from the subject *and her mental life*," we could plausibly read Frise as invoking the following principle as a key premise for the instability problem:

MJ Necessarily, any event that is *mentally inaccessible* to the subject of target belief B cannot constitute any part of the explanation for B's justificatory status.

³⁰ See Conee and Feldman (1998) for an extended discussion of the generality problem.

³¹ This approach to the generality problem might be attractive to one who thinks that the concept of justification fundamentally plays the role of allowing us to flag reliable informants in our community (on various matters). See Sosa (1991: 275) for a discussion of this "reliable informant" approach to the concepts of justification and knowledge.

³² Thanks to an anonymous referee for highlighting the relevance of modal reliabilism to this version of the instability problem.

Were MJ true, it couldn't be the case that Maria's *the sun is still setting* beliefs are justified at t_1 but then cease to be justified by t_2 due to some event in the interim—of which she's completely unaware—that eliminates her perceptual reliability. Suppose, unbeknownst to her, in the interim her brain is hit by gamma radiation, radically scrambling her cerebral cortex and causing her to have all sorts of vivid hallucinations. From a reliabilist perspective, these sorts of factors would at least partially determine whether Maria's beliefs are epistemically justified. But, insofar as the gamma radiation and her subsequent unreliability are inaccessible to her upon reflection, MJ entails that such events cannot affect the justificatory status of her *the sun is still setting* beliefs.

In response, many reliabilists will insist that mentally inaccessible factors *could*, in principle, alter the justificatory status of our beliefs. This is because, reliabilism, as it's traditionally construed, is an externalist theory of justification. According to externalist theories, factors that are either external to the cognitive agent or cognitively inaccessible to her can determine whether a belief has justification. Principles like MJ have been at the center of debates pertaining to justification for decades.³³ MJ is highly controversial, and traditionally, externalists have found principles like this to be unmotivated.

That said, to the extent that the instability problem relies on MJ, it would seem to beg the question against the reliabilist. On the other hand, reliabilists have ready responses to versions of the instability problem that don't rely on MJ, as I argued above. In either case, the conclusion of the instability problem remains unmotivated. Therefore, in its current formulation, the instability problem gives us no compelling reason to reject reliabilism as specified with the proximate reliability parameter.³⁴

8. Conclusion

Giving a principled and thorough account of the temporal parameters governing reliability is certainly an important task for the reliabilist to complete. Just like providing an informative answer to the generality problem, offering an account of these parameters would increase the explanatory power of reliabilism. However, as we've seen, Matthew Frise thinks that the

³³ For an influential defense of a mental accessibility requirement on justificatory factors, see Ginet (1975: 34-36). For arguments against such a requirement, see Alston (1986: 203-15). There, Alston shows that the extant arguments in favor of a mental accessibility requirement on justificatory factors either presuppose the controversial deontological conception of justification, or invoke principles that make the requirements for justification far too demanding (e.g., requiring that a belief have second-order justification just in order to have first-order justification). Also see Bergmann (2006: 13-19) for a similar argument against accessibility or "awareness" requirements on justificatory factors.

³⁴ In his presentation of the instability problem, Frise presents another version of this argument that seems to invoke a premise like P**J**b:

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 (938-9, emphasis mine).

Here, Frise seems to merely re-raise the unnatural deduction problem for dynamic parameters like proximate reliability—to which I have responded in section 3.

temporality problem constitutes a new objection to reliabilism. But the three arguments aimed at demonstrating the temporality problem each have serious flaws. As I've shown above, the reliabilist can invoke non-ad hoc parameters and other background assumptions to parry the formulations of these arguments that only invoke plausible premises. There are other interpretations of these arguments that leave the reliabilist without a straightforward solution. But these latter interpretations invoke controversial claims that are at best unmotivated in the essay, and at worst depend on substantive theses that merely re-raise old debates about externalism and reliabilism. That said, the temporality problem, in its current state, fails to deliver a new reason to reject reliabilism.

References

- Alston, William P. (1985), "Concepts of Epistemic Justification." *The Monist*. 68.1: 57-89.
- (1986), "Internalism and Externalism in Epistemology." *Philosophical Topics*. 14.1: 179-221.
- (1988), "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives*. 2: 257-299.
- Audi, Robert (2011), *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. 3rd Ed. New York: Routledge.
- Bergmann, Michael. (2006), *Justification Without Awareness*. Oxford: OUP.
- Conee, Earl. (1992), "The Truth Connection." *Philosophy and Phenomenological Research*. 52.3: 657-69.
- Conee, Earl and Richard Feldman. (1998), "The Generality Problem for Reliabilism." *Philosophical Studies*. 89.1: 1-29.
- (2008), "Evidence." *Epistemology: New Essays*. Ed. by Quentin Smith. Oxford: OUP. 83-104.
- Frise, Matthew. (2018), "The Reliability Problem for Reliabilism." *Philosophical Studies*. 175.4: 923-45.
- Ginet, Carl. (1975), *Knowledge, Perception, and Memory*. Dordrecht: D. Reidel.
- Goldman, Alvin I. (1979), "What is Justified Belief?" *Justification and Knowledge*. Ed. by G.S. Pappas. Dordrecht: D. Reidel Publishing. 1-23.

- (1994), “Naturalistic Epistemology and Reliabilism.” *Midwest Studies in Philosophy*. 19.1: 301-320.
- (1999), *Knowledge in a Social World*. Oxford: OUP.
- (2002), “The Unity of the Epistemic Virtues.” *Pathways to Knowledge*. Oxford: OUP. 51-72.
- (2012), “Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism’s Troubles, Reliabilism’s Rescue Package.” *Reliabilism and Contemporary Epistemology*. Oxford: OUP. 123-50.
- McCain, Kevin. (2014), *Evidentialism and Epistemic Justification*. New York: Routledge.
- Sosa, Ernest. (1991), *Knowledge in Perspective*. Cambridge: CUP.