UPPING THE EX ANTE PROBLEM FOR RELIABILISM

Matthew Frise

Penultimate draft

Abstract: Process reliabilism is a theory about *ex post* justification, the justification of a doxastic attitude one has, such as belief. It says roughly that a justified belief is a belief formed by a reliable process. It is not a theory about *ex ante* justification, one's justification for having a particular attitude toward a proposition, an attitude one might lack. But many reliabilists supplement their theory such that it explains ex ante justification in terms of reliable processes. In this paper I argue that the main way reliabilists supplement their theory fails. In the absence of an alternative, reliabilism does not account for ex ante justification.

Process reliabilism is a theory about epistemically justified belief. It is roughly the theory that a justified belief is a belief formed by a reliable process. Reliabilists often call the justification of a doxastic attitude like belief *ex post* (or *doxastic*) justification. But that is not the only sort of epistemic justification. In some cases, a subject is justified in believing that *p* and yet she does not believe that *p*. And in some cases, she is justified in believing that *p*, and she does believe that *p*, but her belief is improperly formed—perhaps because it was not caused by something that confers justification. These cases highlight that not all justification belongs to the doxastic attitude one in fact has. Reliabilists often call this other justification *ex ante* (or *propositional*).

As a theory of justified belief alone, reliabilism does not thereby provide an account of ex ante justification. Having ex ante justification does not require forming a belief, much less doing so reliably. But reliabilism can be supplemented in such a way that it does offer a theory for ex ante justification, a theory that centers on reliable processes. Reliabilism, consequently, explains ex ante justification in terms of ex post justification. This order of explanation is among reliabilism's most

provocative innovations, and may have striking results (Kornblith 2017: 64-6; Neta 2017). Suppose ex post justification importantly depends on our psychology. Its being more fundamental than ex ante justification would then imply that the latter also importantly depends on our psychology. Not even ex ante justification would be entirely theorizable a priori.

In this paper I argue that the main way that reliabilists supplement their theory, such that it accounts for ex ante justification, fails. If reliabilism cannot account for ex ante justification, then there is less reason to suppose that ex post justification is fundamental and ex ante justification is derivative. A more important upshot of my argument is that reliabilism faces a problem. Since ex ante and ex post justification are so tightly linked, the proper theory of each should be linkable (cf. Goldman 1979: 22). Reliabilism is deficient if it would estrange these two kinds of justification. Reliabilism is also deficient if the correct theory of ex ante justification does not concern reliable processes. At the end of the paper I explore two ways reliabilists might avoid this upshot, and their costs.

Alvin Goldman (1979: 21) offers the original reliabilist account of ex ante justification. He says a subject is ex ante justified in believing p at t just when 'a reliable belief-forming operation is available to him such that the application of that operation to his total cognitive state at t would result, more or less immediately, in his believing p and this belief would be *ex post* justified.' Being justified in believing that p consists, roughly, in having available a reliable process that would produce a justified belief that p, if that process were active. More precisely:

> *Ex Ante Justification* (EAJ). S is justified in believing p at t *iff* there is a reliable belief-forming operation r available to S such that S applies r to S's total cognitive state at t counterfactually implies that at $t+n$ S has a justified belief that p.

Here, n must be 'suitably small'. EAJ explains ex ante justification in terms of a counterfactual justified belief, and reliabilism theorizes about that belief's justification. So, given EAJ, reliabilism

appears to account for ex ante justification. EAJ is largely unchallenged, and alternatives in the literature do not relevantly differ.[1] Goldman (1979: 22) says EAJ 'falls out directly' from reliabilism. But even if it does not, EAJ does make sense of many cases, and that helps explain its appeal. Maria has not yet formed the belief that she is excited. But when she stops to try, she tends to accurately introspect her mood; a reliable introspective process is currently available to her. If she *were* to introspect, she would indeed justifiedly believe that she is excited. Prior to introspecting, however, it's plausible that for her, believing that she is excited is justified. EAJ rightly says she has this justification. After all, if she paused to reflect on how she's feeling, she'd justifiedly believe she is excited.

Most reliabilists understand the *reliability* of a process in terms of its ratio of true-to-false outputs in a domain of worlds (Goldman 1986; Henderson and Horgan 2010). That is, in a special set of worlds, the process produces a sufficient number of true beliefs relative to false beliefs. What this ratio is, exactly, is underexplored. Goldman (1986: 103) thinks it is better than 1-to-1. It matters little for my purposes, so I will just call it *good enough*. It also matters little which domain of worlds we specify as determining a process's reliability. Here is why EAJ is no good.[2]

Suppose, in one of these worlds, that belief-forming operation r is available to S. Suppose r's ratio of true-to-false outputs is good enough, but only just. So, r is reliable. Suppose that if S were to apply r to her total cognitive state at t, then S would believe p at $t+n$. And suppose r is the only available belief-forming operation that would lead to this belief. Finally, suppose p is false.

Because p is false, if S were to believe p at $t+n$ by process r, r would at $t+n$ not be good enough. At $t+n$, r would have produced one too many false beliefs. At $t+n$ r would not be reliable.

[1] Cf. Henderson and Horgan (2010: 201); Lyons (2016: 167): 'We should say that S is ex ante justified in believing that p at t if and only if, if S were to come to believe p using only processes that took as inputs mental states S is in at t, then S would be (doxastically) justified in believing p'; Senor (2017: 331): 'Similarly, one has *ex ante* justification for a proposition just in case this conditional is true: if one were to come to believe the proposition, one would be justified in his belief.'
[2] That is, EAJ is no good on the accounts on which reliability is settled by a performance in a domain of worlds. An underdeveloped alternative interpretation of reliability is that it is a propensity, something with no modal measure. But EAJ, as a counterfactual theory of ex ante justification, would pair oddly with this alternative.

So, at $t+n$ S's belief that p would not be the output of a reliable belief-forming process. Reliabilism therefore implies that this belief would not be justified. Because this belief would not be justified, the right half of EAJ is not satisfied; r is the only available process by which S would form the belief that p, and satisfying the right half of EAJ requires S to have available a process that would lead to *justified* belief. Since the right half of EAJ is not satisfied, EAJ implies that S is not justified in believing p at t. And yet r, the process by which S would form belief that p, is reliable at t.[3]

More concretely: suppose Sara is training to interpret complex readouts of an instrument that measures midichlorian quantities in a chamber. She stares at the instrument. The only available belief-forming operation available to her for interpreting the readout has been only marginally good enough so far. That process, r, is therefore reliable. If she were to apply r now, however, she would misinterpret the panel, forming a moment later the false belief that there are over 27,000 midichlorians in the chamber. As a result, r would have produced one too many false beliefs, and would not be reliable. So, the belief would not be the output of a reliable belief-forming process. It would not be justified. So, EAJ's right half is not satisfied. EAJ, consequently, implies that Sara is not now justified in believing that there are over 27,000 midichlorians in the chamber. Still, right now, since Sara has not formed the false belief, r is reliable.

Before I explain why EAJ's implication here is incorrect, I will make clear how EAJ could even have this implication, given that r is reliable at t. On a theory of reliability, reliability is either *dynamic* or *static*—it is either capable of changing within a world over time, or incapable (Frise 2018). Suppose reliability is dynamic. Suppose, for example, reliability can change over time as a process's track record changes. Within a world w, a process could be reliable at t, but unreliable at $t+n$, if it

---

[3] It doesn't matter whether the truth-ratio good enough for reliability is vague. The problem I am articulating still arises, since one additional false belief can render a formation process not determinately good enough, and therefore not determinately justifying. So at t, a process could determinately provide ex post justification but not ex ante justification. Also, even if reliability is vague, the problem still arises on a tweaked case: suppose at $t+n$ r would form not just belief that p but also a number of closely related false beliefs, and no true beliefs. At $t+n$ r would not be reliable, and this assumes no precise cutoff between reliability and unreliability.

forms too many false beliefs by $t+n$. If that happens, a belief formed by that process at $t+n$ is unjustified. So, even though the process is reliable at t, the right half of EAJ would not be satisfied at t. (When assigning ex ante justification at t, EAJ looks at whether there would be justified belief at $t+n$ if r were applied at t. Suppose Sara applies r now in our world, and in the next moment forms the belief that there are over 27,000 midichlorians in the chamber. The falsity of this belief renders r unreliable a moment from now, and so the belief is not justified then. So, EAJ is not satisfied now.) A process at t is providing ex post yet not ex ante justification.

Now suppose instead that reliability is static—whether a process is reliable never changes within a world. Consider a version of reliabilism sometimes called *Same World Reliabilism*: a subject's belief is justified in a world just in case it is formed by a process that is reliable in *that* world. It could be that in S's world, w, S does not apply process r at t, and r otherwise maintains a good enough ratio of outputs. So, at all times in w, r is reliable. Call w^* the nearest world in which S does apply r at t. Applying r at t results in belief that p at $t+n$. And p is false. A static account of reliability can imply that, since r's ratio of true-to-false outputs in w^* is not good enough at $t+n$, r is unreliable at all times in w^*. Belief formed in w^* at any time by r is unjustified. As a result, the right half of EAJ is once again not satisfied; applying r to S's total cognitive state at t counterfactually implies that at $t+n$ S has an *unjustified* belief that p. This is because this counterfactual looks to the nearest world in which S applies r, and that world is w^*, where r is unreliable. And this limits ex ante justification in w. (Sara does not in fact apply r and go on to form the false belief that there are over 27,000 midichlorians in the chamber. But in the nearest world in which she does apply r, it is and has always been unreliable in that world, because in that world it yields marginally too many false beliefs. Since applying r in

that world leads to an unjustified belief, EAJ is not satisfied in our world.) Yet, because r is reliable in w, it is a source of ex post justification there.[4]

For at least two reasons EAJ's implication is wrong. First, it incorrectly disassociates ex post and ex ante justification. Whenever one has ex post justification, one should simultaneously have ex ante justification. What's more, part of whatever explains the ex post justification should help explain the ex ante justification. If, say, the perceptual process that forms my belief that q justifies it now, something about that process should also justify for me believing that q now. Similarly, if the perceptual process that forms my belief that q justifies it now, something about this process should furnish me now with ex ante justification for believing a wide range of other propositions. Applying this process should lead to many additional justified beliefs. EAJ's implication conflicts with this. At t, r is reliable. So the beliefs it has formed by t have ex post justification. Yet r is not at t a source of ex ante justification for believing p.

Here is a second reason that EAJ's implication is wrong. On EAJ, in some cases, whether one has ex ante justification for believing p directly depends on whether p is *true*, for *any p*. The falsity of p would compromise the reliability of a process that would otherwise be only just good enough and that forms belief that p. The truth of p, however, would allow that process to remain reliable. Because reliability depends on truth in this way, on EAJ, ex ante justification depends on truth. Perhaps believing some false propositions, given their specific content, cannot be ex ante justified. But the examples in this paper do not depend on p's content. EAJ would incorrectly make ex ante justification covary with truth for just any proposition, when the conditions I outlined are met. And

---

[4] It might seem that r's reliability should remain fixed when determining which worlds are nearby, and so r cannot be unreliable in a nearby world at $t+n$ if it is reliable here at t. But a process's reliability supervenes on its performance. Massive similarity in supervenience bases can take precedence over potential supervenient properties (such as reliability), when ordering worlds. So, it's possible that a process's reliability differs in a nearby world, if the events that determine reliability in that world are similar enough.

the conditions I outlined can be easily met. There are countless processes. Many will have a ratio of true-to-false outputs that is only just good enough.

Given these problems, reliabilists have three main options, options that are not mutually exclusive. Critics of reliabilism will view each option as costly, and might therefore view this as a trilemma. Reliabilists might instead see the options as refinements of their theory. The first, and perhaps most obvious and least revisionary option, is to reject EAJ. (If Goldman is right that EAJ 'falls out directly' from reliabilism, then reliabilists might however consider rejecting reliabilism too.) Since the main alternatives to EAJ do not relevantly differ, their perils are similar. So, reliabilism faces what we might call the *ex ante problem*, the problem of plausibly accounting for ex ante justification. Without a solution to it, reliabilism is a defective theory of justification, severing ex post from ex ante justification and also failing to make reliability central to ex ante justification.

A second main option for reliabilists in light of EAJ's problems is to reject the general type of account of reliability that reliabilists tend to accept. On this kind of account, a process's performance in a certain domain of worlds determines whether the process is reliable. If the process's performance in the domain is good enough, the process is reliable; if not, the process is unreliable. One might reject this account and instead understand reliability as a primitive propensity, such that a process could be reliable or unreliable regardless of how it performs in any domain of worlds.[5] Reliability, then, would not be analyzable in terms of counterfactuals and would not ensure the truth of many intuitive counterfactuals concerning performance; that a process is reliable in a world or domain of worlds may not counterfactually imply that it performs well enough there.

This propensity interpretation of reliability is little explored, but it is nonetheless available to reliabilists. And the problem I raised for EAJ does not seem to arise on this propensity account of reliability. The formation of an additional false belief would not compromise a process's reliability,

---

[5] See Frise (2018) for further discussion of theories of reliability. I thank an anonymous reviewer for inviting me to address this second main option further.

because a process's reliability is not determined by its performance. A marginally reliable process could produce another false belief, yet remain reliable. Applying a marginally reliable process could result in another false and yet justified belief. So, in cases like Sara's, the right half of EAJ can be satisfied. In those sorts of cases a marginally reliable process can be not only a source of ex post justification, but also ex ante justification. The ex ante problem appears easily solved.

If this paper merely provides reason to adopt a less explored account of reliability, its outcome is still noteworthy. However, I think the ex ante problem still arises on this propensity account of reliability. It arises not because EAJ has problematic implications on it, but because EAJ fits poorly with it.

EAJ fits poorly with a propensity account of reliability because EAJ essentially relies on a counterfactual. EAJ says a subject has ex ante justification just in case there is available to her a reliable process such that the subject applying the process *counterfactually implies* that she has a justified belief shortly after. Yet the main reason to reject a performance-based account of reliability and to accept instead a primitive propensity account has to do with spurning modal reductions in our epistemic theorizing. A primitive propensity account is attractive because, roughly, performance in a domain of worlds seems like the wrong sort of relation to focus on when reconnoitering what matters for justification. Yet EAJ would have us focus on a relevantly similar relation when assessing ex ante justification; in a relevant way, EAJ is a performance-based account of ex ante justification. EAJ measures ex ante justification with the availability of a process that performs well in a domain of worlds. In the nearest world where the process is applied, it yields a justified belief. It is plausible that whatever supports a propensity account of reliability should support a *propensity account of ex ante justification*, rather than EAJ. A propensity account of ex ante justification would look at the propensity of a process, not at what the process does in some domain of worlds, when attributing ex ante justification. But just what is a precise statement of such an account? It remains to be seen. EAJ

is all we have so far. In the absence of a propensity account of ex ante justification, the ex ante problem remains. Reliabilists need to plausibly account for ex ante justification.

The third main option for reliabilists, given EAJ's troubles, is perhaps the most revisionary. This option aims to sever reliabilism from ex ante justification, apparently eliminating reliabilism's need to account for ex ante justification. One could support the severance in a few ways. One could show there is no reason to suppose there is such a thing as ex ante justification. Or one could show ex post justification has no close connection with ex ante justification. Either way, reliabilism owes no account of ex ante justification. There is no ex ante problem.[6]

This third option cuts a big tie with traditional reliabilism. Since reliabilism's inception, reliabilists like Goldman (1979: 21) have understood their theory to be 'closely related' to an account of ex ante justification (which they have taken to exist!). Why sever? After all, at first glance there appears to be such a thing as ex ante justification, and it appears closely connected to ex post justification. To illustrate this, Goldman (1979: 21) gives the case of a subject who does not believe that p, and who asks us whether she should come to believe that p. We reply that she is justified in believing that p. Since there is no belief here, this cannot be ex post justification. It seems instead to be ex ante justification. And ex ante justification appears related to ex post justification. Having the latter at least appears sufficient for having the former, and both are evaluative features apparently necessary for knowledge. If appearances here are correct, reliabilism must account for ex ante justification. We would need reason to reject appearances here—otherwise, severing reliabilism from ex ante justification would be a mere ad hoc escape from the ex ante problem.

Additionally, even if the severance were made plausible, reliabilism would not yet be out of trouble. If the severance is because there is no such thing as ex ante justification, then EAJ seems false. Even if there is no such thing as ex ante justification, the right half of EAJ can still be satisfied

---

[6] I thank an anonymous reviewer for mentioning this third main option.

simply by the right sort of process being available to a subject. If the right half is satisfied, EAJ implies a subject has ex ante justification. This implication is incorrect if there is no ex ante justification. So EAJ is incorrect. Yet, as noted earlier, Goldman (1979: 22) claims that EAJ 'falls out directly' from reliabilism. If that means reliabilism implies EAJ, then the falsity of EAJ results in the falsity of reliabilism. If, however, the severance is because ex post justification has no close connection with ex ante justification, then reliabilists need to explain why EAJ, a theory of ex ante justification, still falls out directly from reliabilism, a theory of ex post justification. Neither way of severing reliabilism from ex ante justification is clean.

I am inclined to accept the first main option: EAJ is false, and so reliabilism faces the ex ante problem. The source of the problem is that EAJ is a performance-based, counterfactual account of ex ante justification. It allows a subject's ex ante justification at a time to depend on what she would go on to believe. But things might play out badly. It could be that an important property would thereby no longer obtain, with unintuitive implications for the relation between ex post and ex ante justification. The ex ante problem should arise for other theories of ex post justification that rely on performance-based, counterfactual accounts of ex ante justification. It might not arise for reliabilists who can motivate other sorts of accounts of ex ante justification. I suspect that the ex ante problem arises for any theory that understands ex post rather than ex ante justification to be fundamental, but confirming this is tomorrow's work. Reliabilists may prefer one of the other main options. Any would notably advance theorizing about reliabilism.[7]

Department of Philosophy

Santa Clara University

References

Frise, M. (2018). 'The Reliability Problem for Reliabilism.' *Philosophical Studies* 175, pp. 923–945.

Goldman, A. (1979). 'What Is Justified Belief?' in G. Pappas (ed) *Justification and Knowledge*. (Boston: D. Reidel), pp. 1–23.

Goldman, A. (1986). *Epistemology and Cognition*. (Cambridge: Harvard University Press).

Henderson, D. and Horgan, T. 2011. *The Epistemological Spectrum: At the Interface of Cognitive Science and Conceptual Analysis*. (Oxford: Oxford University Press).

Kornblith, H. (2017). 'Doxastic Justification Is Fundamental.' *Philosophical Topics* 45, pp. 63–80.

Lyons, J. (2016). 'Goldman on Evidence and Reliability,' in H. Kornblith and B. McLaughlin (eds) *Goldman and His Critics*. (Malden: Wiley), pp. 149–174.

Neta, R. (2017). 'Two Legacies of Goldman's Epistemology.' *Philosophical Topics* 45, pp. 121–36.

Senor, T. (2017). 'Preservation and Generation,' in S. Bernecker and K. Michaelian (eds) *The Routledge Handbook of Philosophy of Memory*. (New York: Routledge), pp. 323–334.