



ARTICLE

A Reliabilism without Counterexamples and Reference Class Problems

Balder Edmund Ask Zaar 

Department of Philosophy, Lund University, Lund, Sweden
Email: balder.ask_zaar@fil.lu.se

(Received 10 December 2024; revised 21 March 2025; accepted 13 May 2025)

Abstract

In this article, I present a modified reliabilist theory of knowledge which purports to solve many of the problems currently facing standard reliabilism. With the help of a dispositionally construed reliabilist theory of knowledge and justification (DRK and DRJ), tentative responses to the following problems for reliabilism are offered: The New Evil Demon Problem, The Clairvoyance Problem, The Mr. Truetemp Problem, The Gettier Problem, Barn Cases (Brandom's and Goldman's), and The Lottery Problem. Lastly, I argue that, despite diverging from the letter of standard reliabilism, DRK and DRJ remain within the spirit of reliabilist epistemology

Keywords: Reliabilism; dispositions; dispositional reliabilism; propensity; solutions

1. Introduction

Standard process reliabilism states that a person has knowledge if and only if they acquire a belief by way of a reliable belief-forming process. “Reliable” is typically understood in a frequentist sense, where the reliability of a belief-forming process has to do with producing a high ratio of true beliefs. This form of reliabilism faces many purported problems. Each one engenders a *reductio ad absurdum* argument against reliabilism. This generally happens along the following lines: One assumes the truth of reliabilism, either $RTB \equiv K$ and $R \equiv J$ (reliable true beliefs are equivalent to knowledge, and belief acquired with reliable processes are equivalent to justification), and derive a contradiction, which then entails the negation of the initial assumption ($\neg (RTB \equiv K)$ or $\neg (R \equiv J)$, or both). Among the more well known that bring about a *reductio* argument for reliabilism, we find: The New Evil Demon Problem (Cohen and Lehrer 1983; Cohen 1984), The Clairvoyance Problem (BonJour 1980), The Mr. Truetemp Problem (Lehrer 1990), the Gettier Problem (Gettier 1963), Goldman's (1976) and Brandom's (1998) barn cases,¹ and The Lottery Problem (Adler 2005).

¹These I will not present as *reductio* arguments, because I view them, not exactly as counterexamples, but as explanatory challenges. Although one could, of course, with a high degree of contrivance, present them as *reductio*-arguments.

To my knowledge, there is no theory of reliabilism that purports to solve all of these problems at once. In Goldman (1986), for instance, we find responses to the generality problem and the problem of clairvoyance and a version of the new evil demon problem. Comesaña (2002, 2006, 2009) has an excellent track record of solving problems for reliabilism with a number of different strategies. Greco's (2010) tackles a range of classical Gettier-style counterexamples, as well as the generality problem and the value problem (Kvanvig 2003; 2010). Sosa's (1992) virtue reliabilism solves the new evil demon problem and the clairvoyance problem, but by bifurcating the concept of justification into weak and strong justification (apt and adroit).² One of the more impressive attempts (that seem largely successful), and also most similar in aim to my own, is Pritchard's *Anti-Luck Virtue Epistemology* (2012). What this paper adds to this literature is (1) a distinct version of reliabilism capable of achieving much, if not all, that the theories above achieve without a pluralism about justification and/or knowledge, and without adding normative notions to reliabilism (i.e., virtue reliabilism) and (2) an argument for why the analysis is reliabilist in spirit.

I believe reliabilism contains solutions to the problems listed above, but that some modifications to our understanding of reliabilism is required in order to do so. In this article, I will attempt to supply one such understanding that I will then apply to each of the problems above and argue that they are solved upon adopting the relevant analysis of reliabilist knowledge and justification.³

The aim of the paper is thus twofold: (1) give a set of individually necessary and jointly sufficient conditions for knowledge that avoids or solves – in one way or another – the above-mentioned problems for reliabilism, (2) while maintaining the spirit of reliabilist epistemology. A subordinate goal with this analysis of reliabilist knowledge is to relate reliabilism back to its aetiological origins, as a causal theory of knowing, but by understanding the epistemically relevant causal relationship as dispositions that require for their manifestation the cooperation between the external world and subject-internal cognitive mechanisms.

I will proceed as follows. In section (2), I will present the above-mentioned analysis of reliabilist knowledge that I have in mind. In section (3), I will show how the theory presented in (2) avoids being undermined by the types of counterexamples and problems that standard reliabilism typically fails to handle. In section (4), I will argue that my analysis is in fact a reliabilist theory of knowledge, despite the rather drastic changes to the formulation of it.

2. Introducing dispositional reliabilism

I assume that reliable processes are essential to knowledge *and* justification. I will therefore maintain the general structure of the classical JTB analysis. Where I depart from the standard conception of reliabilist knowledge (a true belief acquired by way of a reliable belief-forming process) is that I take the probabilistic notion inherent in reliable relations (that is, the truth conduciveness) to be dispositional. The frequentist understanding of reliability can be said to arise out of a dispositional understanding of reliability.

²For an account of weak and strong justification, see Goldman (1988).

³It is worth noting at this point that my approach may bear strong resemblance to agent reliabilism (Greco 1999, 2010), virtue reliabilism (Sosa 1992), and perhaps most of all with anti-luck virtue epistemology (Pritchard 2012) and what Becker calls “modalized process reliabilism” (Becker 2006). There are relevant distinctions to be made here, but would require too much space for an already lengthy article.

The counterexamples listed in the introduction produce three primary intuitions (as far as I can tell): (1) one can have reliability without being cognitively responsible (justified); (2) one can lack reliability yet still be blameless (justified); and (3) one cannot be justified without having privileged access to whatever it is that makes one justified. These intuitions I call “non-naturalist internalist intuitions” (internalist intuition for short), they stem from viewing justification, and therefore knowledge, as something inherently related to normative notions such as blame and responsibility (in the first two cases), and from the idea that one does not know unless one knows that one knows (in the last case). Reliabilism, formulated *merely* in terms of the conditional probability of acquiring a true belief given the use of a reliable process, cannot explain the cases where our sense of epistemic responsibility and blamelessness converges with our intuitions surrounding epistemic justification. Since reliabilism is a form of externalism, it also trivially fails to account for Lehrer’s claim that knowledge is not merely about having correct information. In other words, according to Lehrer, having reliable access to information is not sufficient to count as a way of knowing. As he writes (1990, 163): “[The Problem] is that more than the possession of correct information is required for knowledge. One must have some way of knowing that the information is correct.” The question for the reliabilist who takes these counterexamples seriously, and in the process approaches the internalist with some hope of reconciliation, is how to meet internalist intuitions without himself becoming an internalist.

For obvious reasons, the externalist-reliabilist cannot merely concede that knowledge inherently has to do with blame and responsibility and a privileged access to one’s epistemic status. To do so, would be to become an internalist. What they *can* do is to provide a theory that does not engender (seemingly valid and sound) *reductio* arguments. But they can also provide a theory that at the very least aligns with the internalist intuitions such that, when the internalist judges “this is not a state of knowledge,” then the same judgment should be derivable from our analysis of knowledge. If an externalist-reliabilist theory of knowledge achieves such an analysis, it would be a significant step toward a reconciliation between externalist and internalist camps. If each of our theories of knowledge capture expert intuitions, it may be less of a concern that our exact wording and degree of satisfaction of non-essential epistemic desiderata differ (cf. Alston 1993).

Reductio-arguments engendered by counterexamples, in my view, are the primary ways in which to show that reliabilist conceptions of knowledge are fundamentally (conceptually) flawed. The first step toward a reliabilist epistemology that “works,” then, is to offer an analysis that does not, along with other plausible premises, lead to contradiction(s). In order to take this step, I will now propose an analysis of knowledge and then demonstrate its capacity to (in most cases) co-vary with the internalist intuitions (1–3 above).

Here is the analysis of knowledge and justification I propose. Let us call them dispositional-reliabilist knowledge (DRK) and dispositional-reliabilist justification (DRJ). A subject *S* knows a proposition *p* if and only if:

- (A) *S* believes that *p*.
- (B) *p* is true.
- (C) *S* has acquired the belief that *p* with a reliable process *R*, where reliability is understood in terms of truth-conduciveness (a high ratio of true beliefs), which in turn is entailed by *R* consisting of:
 - I. *S* using a cognitive process type CPT disposed to produce true beliefs when acquiring the belief that *p*.
 - II. *S* using a token of the CPT in (for the purposes of obtaining true beliefs) safe circumstances.

Since (C) is entailed by I. and II, one could say also say that S knows that p if and only if:

- (A) S believes that p .
- (B) p is true.
- (C) S uses a cognitive process type CPT disposed to produce true beliefs to acquire the belief that p and S using a token of the CPT in safe circumstances.

The following are the necessary and sufficient conditions for *justification*. S is justified in believing that p if and only if:

- (A_j) S believes that p .
- (B_j) S uses a cognitive process type CPT disposed to produce true beliefs to acquire the belief that p .

The disposition in question can be understood (merely as a heuristic, or in a highly idealized way, as I will make clear in section 4) in terms of a conditional analysis:

(CA): a CPT has the dispositional property F to produce a true belief that p given that the user S of CPT is in a position to acquire a true belief if and only if, were S to use a token of the CPT, then S would produce a true belief that p .

Here I take a cognitive process type to, roughly, involve our perceptual faculties, but also various forms of extended cognition, which allows for knowledge acquisition with the help of various instruments (cf. Goldman's distinction between belief-forming processes and belief-forming methods in his 1988). I also take a priori cognition (mathematical and logical reasoning) to be dispositional, with the main difference being that the inputs differ significantly as compared with perceptual processes. A priori reasoning, on my view, involves a disposition that takes something like a thought (or the apprehension of a proposition serving as a premise in an argument) or a set of thoughts as its input, and outputs a true belief. The exact details of what different types of knowledge-acquiring cognitive processes would be, I think is ultimately an empirical question. A CPT here, then, is merely intended as a placeholder for some empirically explicable concept.

I want to sketch how I understand "being in a position to acquire a true belief". This locution is meant to broadly capture the types of situations wherein one acquires knowledge. Essentially, situations such as when one interfaces with the external world via perception, or when reading thermometers to get information about the temperature. Roughly, if the world "discloses" itself (Merleau-Ponty 2014) either directly via perception or indirectly via scientific instruments, then one is in a position to acquire a true belief, and one will also do so if the cognitive mechanisms one uses to apprehend the world have the dispositional property to produce a true belief.

Safety plays a crucial part in my analysis. I roughly understand it in the sense of "could not easily have been wrong" (cf. Williamson 2000; Sosa 1999). Safety appeases our anti-luck intuition, and thus serves to rule out cases of luckily true beliefs as being cases of knowledge. Here is a slightly more precise version of safety:

(SAFETY): An epistemic subject S acquires a belief that p via some CPT in epistemically safe conditions if and only if S could not *easily* have formed a false belief via the use of the CPT, given similar circumstances and stimuli.

Safety is about a subject's ability to continue to form true beliefs in similar circumstances, given similar stimuli (Pritchard 2012, 256–257; Unger 1968, 159–160) and methods. One could understand this as, given one's use of a CPT, then, in near-by worlds with the same general circumstances and stimuli, then one is not going to form false beliefs using the same CPT. Note that this does not necessitate that one forms *the same belief* in every case of similar conditions and stimuli.

For the upcoming two sections, it will be also helpful to briefly explain what mimics and maskers are. In short, they are notions that arose from attempts to provide counterexamples to the conditional analysis of dispositions. Mimics and maskers illustrate the idea that the notion of a disposition allows for there to be cases in which there is a disposition at play, but it fails to manifest the type of events that are typically associated with the disposition, but also that there can be cases that seem to have to do with an object's disposition, but there really is not one at play.

Mimics are things that make it seem like there is a dispositional property instantiated because of a constant conjunction between two events, but where this is merely a contingent fact about the object, and not a result of a dispositional property of the object. An example of this would be if some demon was to make every belief acquired by way of clairvoyance true for some subject S. This would mimic a disposition to produce true beliefs in S, but we would not want to say that there is a genuine dispositional property to produce true beliefs involved, since the manifestation of true beliefs does not come about through the right kind of disposition. It would be incorrect to ascribe knowledge to such a subject because their beliefs would not be true in virtue of their dispositional powers, but someone else's. Maskers are things that mask the manifestation of a dispositional property. If one is in rainy conditions, then it may be the case that one is unable to light one's matches. But it nonetheless is the case that the match has the disposition to light when struck. An epistemic example would be something like: if through some evil experiment, one would link the opening of S's eyes to some elaborate light-blocking device that completely blocks out all light in immediate surroundings of S, then S would not be able to have true perceptual beliefs about objects in their surroundings. Nonetheless they would have a cognitive mechanism (a visual system) disposed to produce true beliefs. It is merely being masked.⁴ I want to make clear too that dispositions here are interpreted in a realist sense, but that specifying exactly what kind of internal mechanism is involved in dispositions has to be kept unspecified. The ideal conditions for the flawless manifestation of a disposition may not be specified with ultimate precision, but the stimulus and manifestation conditions (here the disclosure of the world and the resultant true belief, respectively) must be seen as being connected by way of some physical causal process.⁵

With the key notions explained, we can proceed. Section 2.1 will show how my analysis avoids Cohen's New Evil Demon Problem by way of (A_j), in 2.2 I will show how Bonjour's Clairvoyance Problem is solved by way of (A_j), 2.3 deals with Lehrer's Mr. Truetemp via argumentation that does not refer to the definitions above, 2.4 the Gettier problems by way of safety, 2.5 Goldman's and Brandom's Barn Cases by way of safety, and lastly, 2.6, Adler's Lottery Problem by way of characterizing CPTs in a way that excludes lotteries as counting as appropriate ways of acquiring knowledge.

⁴For a discussion of maskers and mimics, see Bird (1998; 2020) and Handfield & Bird (2008), also Gundersen (2003; 2010; 2012) for further discussions of maskers, mimics and alters, especially his (2003) discussion of some principles relating to subjunctive conditionals that may let us remove the anti-masking and anti-mimicry clauses from the conditional analysis.

⁵Cf. e.g., Mumford (1998, 2009) realist view about dispositions.

2.1. New evil demon world inhabitants are reliable

The New Evil Demon Problem (Cohen and Lehrer 1983; Cohen 1984) is a counterexample to reliabilism as a theory of justification. It goes like this: Consider an entire world of highly responsible epistemic agents, called NED-worlders.⁶ They have scientific communities that conduct similar research as we do, they have all the ordinary types of perceptual beliefs that we do, but due to the inference of the evil demon, all the methods they use to acquire their beliefs are highly unreliable. Now, do the NED-worlders really have no justified beliefs due to the influence of the evil demon? All of their type-processes are *de facto* unreliable,⁷ and so according to reliabilism, they cannot have any justified beliefs. Yet we must admit that they are highly epistemically responsible (just as responsible as we are, in fact), and that the fact that they have a high ratio of false beliefs is beyond their ken. They cannot control the epistemic situation they are in, and so why should we deem all of their beliefs unjustified? The internalist intuition, then, says that epistemic responsibility is what matters here. Doing the right thing, epistemically speaking, is what should count towards one's state of being justified, not aspects entirely out of one's control such as whether one's processes are reliable. Doing the right thing, then, has to do with what we are able to control, and we only have privileged access to control over our own behaviors, and so epistemic responsibility must be something that is wholly *internal* to the epistemic agent.

Reliabilism, classically construed as a mere probabilistic relationship between the use of a type of process and a true belief, fails to capture this intuition. As a result, the following *reductio*-argument can be constructed to demonstrate the contradiction that this intuition and reliabilism jointly entail:

- (1) A belief is justified if and only if it has been formed by way of a reliable process (reliabilist definition of justification).
- (2) There exists no reliably formed beliefs in the NED-world (stipulation of the NED-problem).
- (3) There exists justified beliefs in the NED-world. (Internalist Intuition)
- (4) There exists reliably formed beliefs in the NED-world and there does not exist reliably formed beliefs in the NED-world. (\perp via 2, 3 and substitution of justification with reliably formed belief via 1).

Luckily, all we need to do in order to see that justificatory reliabilism – construed as involving the use of a CPT with a certain kind of disposition – does not entail a contradiction is to input the definition of justification given in section 2. The same contradiction simply does not follow:

- (1') A belief is justified if and only if S uses a cognitive process type CPT disposed to produce true beliefs to acquire the belief.
- (2') There exists no reliably formed belief in the NED-world.
- (3') There exists justified beliefs in the NED-world.

⁶NED = New Evil Demon.

⁷Note that we have to interpret the NED-worlders as being unreliable in a frequentist sense. Otherwise, we would be liable to contradict the 'ability intuition' (Pritchard 2012), the intuition that knowledge has something to do with one's cognitive abilities. If make the claim that the NED-worlders are not using dispositions, for instance, it seems odd to ascribe justification to them, in that they would simply not have the ability to acquire true beliefs anymore. I take both the New Evil Demon problem and the Clairvoyance problem as attacking the frequentist understanding of reliabilism.

(4') There exists beliefs acquired by way of cognitive process types disposed to produce true beliefs and there does not exist beliefs acquired by way of reliable processes.

Recall that reliability, understood in a frequentist sense,⁸ does not arise out of the use of a cognitive process alone. One requires something like epistemically safe conditions. The NED-worlders, by stipulation, do not use their cognitive process types in safe conditions. Therefore, they do not have knowledge, and their methods do not yield a high ratio of true beliefs, yet, nonetheless, they are justified in the dispositional sense. Via DRJ, we can account for the fact that the NED-worlders are responsible and why exactly it is that they are doing their best; they are using cognitive processes with the dispositional property to produce true beliefs. If the NED-worlders had better tools available to them (for instance, ones that were reliable in a frequentist sense on top of having the right kind of dispositional property), they would not be responsible if they refused to use them (assuming they are aware of these better tools that were not being affected by the evil demon). The NED-worlders are responsible, and justified in their beliefs, because they are using the only tools available to them, and those tools have the dispositional property to produce true beliefs just like we do. The only problem is that due to the fact that the safety condition on knowledge is unmet, their disposition can be perpetually masked by the influence of the evil demon.

Adopting DRJ as a theory of reliabilist justification achieves two things. One, it shows that the NED-scenario fails to engender a reductio-argument against dispositional reliabilism. Two, it may explain the internalist intuition that what constitutes epistemic responsibility is about using the right kinds of cognitive tools – the ones within the privileged control of the epistemic agent. Moreover, dispositional reliabilism remains externalist, insofar as whether a cognitive mechanism has a certain dispositional property or not need not be in principle accessible to the subject using that mechanism in order for it to serve its purpose (which is to produce true beliefs).

Since the main goal here is to avoid contradiction, and this goal has been achieved with reference to clause (C) from section 2, let us move on to BonJour's Clairvoyance Problem.

2.2. Norman's Clairvoyance – the wrong kind of reliability

BonJour's case is a reversal of the New Evil Demon Problem. Here is the full quote from BonJour (1980, 62):

Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable.

Is Norman, in virtue of his process being reliable, justified in the belief that the President is in New York City? No! He has no reason to think that his clairvoyance is reliable. Again, he would be highly irresponsible if he were to think that his clairvoyance was

⁸"Frequentist sense" here meaning processes that output a high ratio of true beliefs, irrespective of the dispositional properties of the process.

reliable, seeing as there is no available evidence to support the notion that clairvoyance is generally reliable. So again, something is not captured by the reliabilist picture of justification. The frequentist understanding of reliability gives the wrong reading, and it fails to explain why we think that Norman is so irresponsible that he cannot be attributed with a justified belief. The following *reductio* illustrates the effect of this thought experiment:

- (1) A belief is justified if and only if it has been formed by way of a reliable process.
- (2) Norman's belief that the President is in New York City is not justified.
- (3) Norman's belief that the president is in New York City is formed by way of a reliable process.
- (4) Norman's belief that the President is in New York city is formed by way of a reliable process and Norman's belief that the President is in New York City is not formed by way of a reliable process (substituting "justified" in 2 via 1, then the conjunction of 2 and 3 $\models \perp$).

Following the procedure of the previous section, let us simply input DRJ into the argument:

- (1') A belief is justified if and only if S uses a cognitive process type CPT disposed to produce true beliefs to acquire the belief.
- (2') Norman's belief that the President is in New York City is not justified.
- (3') Norman's belief that the president is in New York City is formed by way of a reliable process.
- (4') Norman's belief that the President is in New York city is formed by way of a reliable process and Norman's belief that the President is in New York is *not* formed by way of a cognitive process disposed to produce true beliefs.

Is (4') a contradiction? No. One can yield a high ratio of true beliefs with the help of a cognitive process without it being the case that the cognitive process has the dispositional property to produce true beliefs. In Norman's case, the process is merely frequentist-reliable due to the presence of a disposition mimic. Norman's clairvoyance is *not* a cognitive process disposed to produce true beliefs (it may not even be a cognitive process!). It just so happens that there is a strange set of circumstances that make his clairvoyance look as if it was disposed to produce true beliefs. The resulting true beliefs are thus not formed in virtue of the dispositional properties of his internal cognitive mechanisms. If, on the other hand, there was a world wherein clairvoyance was more or less like perception, where it would be a cognitive process with a dispositional property to produce true beliefs, in virtue of there being some clairvoyance-like sensory system and something like clairvoyance waves, we would have to say that Norman, in fact, has justified beliefs.⁹ But if this were the case, would not the internalist then also have to say

⁹Given that BonJour ascribes a clairvoyant "power" to Norman, we perhaps ought to view Norman as being in a world where clairvoyance-waves and so on are part of the natural world, and that this is what BonJour intended with the case after all. If BonJour's intended point was to say that, even if we were to provide a natural science explanation for the cognitive power, it would be irresponsible in that it was unusual or not sanctioned more generally, then I may have to accept that the argument I make here is somewhat misguided. But luckily, if this is BonJour's intention, then it essentially becomes just like Lehrer's Mr. Truetemp case where, despite there being a natural world explanation for the cognizer's belief, we would not want to attribute knowledge to him in virtue of the fact that his "cognitive" process is too strange and reflectively inaccessible to count as a responsible belief-forming mechanism. I argue against this conclusion

that Norman is in fact epistemically responsible, just as you or I would be if we acquired beliefs by using our visual systems? I think so, and in both cases, DRJ accounts for these intuitions. It is not merely about truth-conduciveness, it is about the right kind of truth-conduciveness. With that said, we can now claim that, again, contradiction has been avoided with reference, again, to (C) from section 2, and so let us move on to the next problem.

2.3. Mr. Truetemp and the internalist encroachment

Lehrer (1990, 162–163) provides the following purported counterexample to externalist theories of knowledge:

Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? Surely not. He has no idea whether he or his thoughts about the temperature are reliable. What he accepts, that the temperature is 104 degrees, is correct, but he does not know that his thought is correct.

Now this thought experiment directly refutes DRK insofar as it is a counterexample against any externalist theory of knowledge. Even if DRJ states that justified beliefs are about subject-internal properties, as long as these are understood as cognitive mechanisms that one does not have privileged access to, then Lehrer's thought experiment does engender a contradiction. And I do take DRJ and DRK to be externalist theories of knowledge and justification. DRJ does not require that one be aware of one's justificatory source in order for one to be justified in holding a belief. One also does not have to know that one knows, in order to know. But this is exactly what Lehrer seems to require of poor Mr. Truetemp. He does not know the temperature, it is said, because he does not know that he has been surgically modified to be able to reliably track the temperature. And so, the *reductio* argument would run as follows:

- (1) K if and only if R.¹⁰
- (2) If one has knowledge, then one has privileged access to one's justificatory source.
- (3) Mr. Truetemp does not have privileged access to his justificatory source.

about Mr. Truetemp in the upcoming section, and if what I say about Mr. Truetemp holds true, then I believe it also holds true for Norman (interpreted as indicated above).

¹⁰“K = R” to read “knowledge is identical with the classical reliabilist conception of knowledge”, as in, a “reliably formed true belief”.

- (4) Mr. Truetemp uses a reliable process to acquire true beliefs (about the temperature).
- (5) Mr. Truetemp has knowledge and does not have knowledge (via 4 + substitution and via 2, 3 and *modus tollens*).

Of course, the externalist has to deny either (2) or deny that Mr. Truetemp actually has knowledge, even given reliabilism. An easy way to do this would be to say that, if we stipulate that Mr. Truetemp is a reflective and rational person, he will not actually believe the contents of his thoughts about the temperature. The issue cannot be resolved this way, however, since Lehrer stipulates that the *tempucomp* forces Mr. Truetemp to accept the contents of his thoughts *unreflectively*. So, Mr. Truetemp has compulsive beliefs about the temperature. Does this not mean that he is entirely epistemically blameless? He surely cannot be blamed for holding these beliefs about the temperature, seeing as he is clearly not responsible for them. They are simply caused by the *tempucomp*. If this is true, then it also seems that the Mr. Truetemp counterexample would work against any theory of knowledge, internalist or externalist, that view justification as having a strong relation to blamelessness. A response to this purported counterexample, then, could be simply to say that it would not only undermine reliabilism, but that it would also undermine any theory of knowledge that attempts to capture the internalist intuition brought about through The New Evil Demon problem – the latter being that epistemic blamelessness is very close to or identical with epistemic justification. If one acquires false beliefs beyond one's ken, and this means that one is epistemically blameless, then there are no reasons to think that just because one acquires true beliefs beyond one's ken, that one would be epistemically blameworthy (and so unjustified). It seems then that, if we were to deny that Mr. Truetemp has knowledge, we would also have to deny that the NED-worlders are justified – for they are equally blameless. The internalist case against the reliabilist, then, is jointly incoherent. But the Mr. Truetemp problem runs a bit deeper than merely being incoherent if taken together with other internalist attacks on reliabilism.

Perception is a paradigmatic way of acquiring knowledge, and our perceptual apparatuses share some similarities with Mr. Truetemp's *tempucomp*. Two primary similarities: (1) we do not choose what we believe on the basis of perception, (2) we do not have access to the complete causal history of our perceptual processes.

When I look out the window at a nearby chestnut tree, do I have any choice in the matter of whether I believe that there is in fact a chestnut tree nearby? Just as Mr. Truetemp with the temperature, no, I do not. The mechanism functions entirely outside of my volition, and the resulting belief, in most cases, does not come by way of an inference like: "I see the chestnut tree. Therefore, I believe that there is a chestnut tree in front of me".

Considering this more generally, would not our unwillingness to ascribe knowledge to Mr. Truetemp imply something rather sinister about how to view our epistemic practices more generally? Has the reader ever looked into the physiological mechanisms that underlie their perception of the world? Does the reader know, were they to look at the sky, where their thought that the sky is blue ultimately comes from? Of course, we do not have access to the full causal history of the types of judgments that we end up believing. We should not need to have such access in order to have a claim to knowledge. Presumably, we still take ourselves to know that we are looking at trees and ducks, and the like, without this type of access. Moreover, if one needs such access in order to be justified in one's more quotidian beliefs (about seeing ducks and trees and the like), it seems that we would invoke an unreasonably high epistemic standard. As a result, knowledge could, and probably would, be unobtainable. Committing to a lower standard

for the obtainment of knowledge, then, seems appropriate as long as we want to maintain that everyday knowledge is obtainable. Does this mean then, according to Lehrer, that I do not have basic perceptual knowledge, just as Mr. Truetemp does not have knowledge about the temperature, just because I do not have access to the causal history of my perceptual faculties? It seems that Lehrer would have to say that I do not. I take this to be an unacceptable all too skepticism-adjacent result. Therefore, if we want to attribute perceptual knowledge to ourselves, we should simply attribute knowledge to Mr. Truetemp.

Given that our paradigmatic cases of knowledge are sufficiently similar to Mr. Truetemp's knowledge, in virtue of his non-volitional acquisition of true beliefs and his lack of access to the causal history of his cognitive process being essentially the same as our paradigmatic ways to acquire knowledge, it seems that the intuition that he does not have knowledge would be too destructive, and so in this case it seems appropriate to simply reject the intuition as either telling us nothing constructive about our concept of knowledge, or as simply highlighting an irreconcilable gap between internalist and externalist conceptions of knowledge and justification.

Lehrer's demands seem not only implausible from the point of view of the externalist, but from any realist about knowledge. We cannot rule out analogous skeptical scenarios in our everyday life and ultimately explain what brings about our beliefs, just as is the case for Mr. Truetemp. If he does not have knowledge, none of us do. And if we do have perceptual knowledge, it also follows that Mr. Truetemp does. The most appropriate thing to do in this scenario is simply to stand with Mr. Truetemp – what goes for him, goes for all of us.

In the following section, I will be discussing some barn cases, one from Goldman (1976), and another more complicated version from Brandom (1998). In both cases, however, I will be able to show more concretely how the safety condition keeps us from attributing knowledge to the people perceiving the only real barn in barn façade county.

2.4. Safety as an antidote to Gettier cases

To spoil it for the reader: safety, again, saves the day. Consider the two Gettier cases, replicated almost completely (1963, 122):

Case I:. Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

(d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

[. . .] Proposition (d) entails:

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d). [. . .] In this case, Smith is clearly justified in believing that (e) is true. But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket.

Case II:. Let us suppose that Smith has strong evidence for the following proposition:

(f) Jones owns a Ford.

Smith's evidence might be that Jones has at all times in the past within Smith's memory owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford. Let us imagine, now, that Smith has another friend,

Brown, of whose whereabouts he is totally ignorant. Smith selects three place-names quite at random, and constructs the following three propositions:

- (g) Either Jones owns a Ford, or Brown is in Boston;
- (h) Either Jones owns a Ford, or Brown is in Barcelona;
- (i) Either Jones owns a Ford, or Brown is in Brest-Litovsk

Each of these propositions is entailed by (f). Imagine that Smith realizes the entailment of each of these propositions he has constructed by (f), and proceeds to accept (g), (h), and (i) on the basis of (f). Smith has correctly inferred (g), (h), and (i) from a proposition for which he has strong evidence. Smith is therefore completely justified in believing each of these three propositions. Smith, of course, has no idea where Brown is. But imagine now that two further conditions hold. First, Jones does not own a Ford, but is at present driving a rented car. And secondly, by the sheerest coincidence, and entirely unknown to Smith, the place mentioned in proposition (h) happens really to be the place where Brown is.

The related argument, as per usual, takes a simple form:

- (1) K if and only if R.
- (2) If a belief is merely accidentally true, then that belief cannot count as knowledge.
- (3) Smith has R (in both cases).
- (4) Smith's belief (in both cases) is accidentally true.
- (5) Smith has knowledge and does not have knowledge.

In order to say that Smith is justified but does nonetheless lack knowledge, we simply have to say that "R" has to do with more than using a certain CPT with the disposition to produce true beliefs. And this is exactly what DRK states. In order to have knowledge, one could not easily have been wrong (that is, the safety condition needs to be fulfilled). In both Gettier cases, we can see that Smith could easily have been wrong. At the same time, Smith can be said to be using a cognitive process type with the disposition to produce true beliefs. He uses reasoning and testimony, for instance, and these are generally considered to be truth-conducive, and so in both cases he is justified, but fails to obtain knowledge.

Now, we may type-individuate what Smith uses to acquire his justified true belief differently, but then it seems that we would change the stipulations of the thought example. For instance, if we say that Smith is actually inferring based on false lemmas, then of course we should not deem him to be justified in his beliefs. But this does not seem to be the level of generality Gettier ascribes to Smith's epistemic process. Smith is using testimony and logical reasoning, and so, understood at that level of generality, he is justified in both cases, given DRJ. If we accept Gettier's stipulation that Smith is using justified methods of acquiring his beliefs, then we need to view them at a level of generality that allows for this reading. When we do that, we can nonetheless see that Smith's use of testimony, memory, and reasoning are not in epistemically safe conditions, and so we get the reading that coheres with our intuitions. Smith lacks knowledge. When Smith thinks that the man who is going to get the job has 10 coins in his pocket, but fails to realize that it is him, he could easily form a belief on the basis of the same method that the man who is going to get the job is Jones, not himself. And so, safety is still lacking for Smith.

2.5. Reference class problems and the safety solution

The barn cases are quite hard to handle on the standard conception of reliabilism. The problem lies in the fact that one has to determine a reference class, that is, the domain

under which a certain probability statement is true, in order to make determinate judgments as to whether a subject in certain confounding scenarios really has knowledge. The question is: what is the relevant cognitive type that “reliable process” refers to. This is the well-known generality problem for reliabilism (Conee and Feldman 1998). For example, when one is looking out at a nearby chestnut tree, does one use visual perception? Or simply *perception*? Or visual perception while wearing glasses? Perception while wearing glasses indoors? Or a process of relying on leaf shapes to form tree-classifying judgments? Each of these process descriptions constitutes different types under which the token process may fall under, and they might all have different degrees of reliability. Without being able to determine the relevant process type, we have no way of saying whether someone is in a state of knowledge or not. The more harmful version of this kind of problem is when it is not only a matter of having different degrees of reliability coupled with each type-individuation, but that same general process can be type-individuated ambiguously, where some individuations yield a reliability judgment whereas others do not. If one is capable of reliably identifying zebras, one may not be seen as maintaining this capacity in a zoo filled with cleverly painted donkeys (an example originating with Dretske 1970, 1016). One may be able to identify barns under normal conditions, but this does not mean that one reliably does so in a county filled with fake barns. Whether one is deemed capable of doing so depends on whether one type-individuates the processes involved as mere “perception,” which is typically reliable, or “perception under deceptive circumstances,” in which case the process would not be reliable. With reference class ambiguity, then, comes ambiguous attributions of justification and knowledge, and so in order for the reliabilist to be able to give principled justification and knowledge attributions in all relevant cases, they need to be able to determine the relevant reference classes for the process involved. With that said, let us get into the scenarios presented by Goldman and Brandom.

Goldman’s scenario runs something like this: Bob is in barn façade county, a place where all barns except one are fake. Bob, epistemically lucky as he is, happens to stumble upon the only real barn. He uses a reliable process to identify this barn, namely visual perception, and so he is completely justified in his belief that there is in fact a barn in front of him. But, of course, we can tell that he merely acquired a true belief by way of luck, and so we should not ascribe knowledge to Bob. Brandom’s example is more complicated in that he wants to evoke a detrimental ambiguity for the reliabilist.

Let us this time stipulate that Bob is technically in Barn Façade County like before. But also, that he is in Real Barn State (wherein Barn Façade County is located), where most of the barns are real. Moreover, he can also be seen as being in barn façade country (wherein Real Barn State is located), where, again, most of the barns are fake. Reliabilism in its classical form offers us no clue as to which reference class (Barn Façade County, Real Barn State, or Fake Barn Country) is the one with which to evaluate the reliability of Bob’s barn-identifying powers.

In both cases we see that Bob lacks knowledge. However, in both cases, we also see that he has a true belief acquired by way of a reliable process, or a process that could be deemed reliable, depending on the reference class under which we evaluate his barn-identifying faculties. But we have not been supplied with a way in which to evaluate Bob’s justificatory status, because we have not been supplied with a way in which to determine suitable reference classes for Bob in the two scenarios. Of course, being able to determine a reference class is a second-order knowledge concern. If we are unable to ascribe knowledge consistently to people in some counterfactual situation, then we cannot say that we know whether they know. But as externalists, we do not have to. In order to respond to the remaining explanatory challenge, however, the safety condition becomes essential.

How does the safety requirement give us a way to evaluate Bob's epistemic status in each situation? It tells us exactly what Bob does not have in both scenarios. If knowledge necessitates safety, and Bob lacks safety, then we have given a conceptual explanation for why Bob lacks knowledge in the two scenarios that coheres with our intuitions. It follows rather trivially that Bob could have easily been wrong, since we have in both cases stipulated that he, by sheer accident, has come in contact with the only real barn in his immediate surroundings. Local safety entails safety simpliciter. Similarly, epistemic "danger" in the sense of "lack of safety", also implies lack of safety simpliciter. Bob, in both cases, is in local epistemic danger (he could easily be wrong), and therefore his state is epistemically unsafe; unsafe simpliciter. Since the safety condition is unmet, then, he also does not per definition have knowledge, and the intuitively correct and unambiguous evaluation of Bob can now be said to be derivable from a reliabilist conception of knowledge.

While the generality problem (Conee and Feldman 1998) is (arguably) another type of reference class problem, and DRK arguably avoids problematic ambiguities in Brandom's barn example, it is unclear whether DRK also solves the generality problem. However, I believe that it may provide some preliminary answers. Obviously, reference to a cognitive process type blatantly runs into the generality problem. Each token process belongs to a potentially infinite number of types. As Conee and Feldman (1998, 3) ask, which process type is it that has to be sufficiently reliable (in the frequentist sense of "reliable")? Now, given DRK, the reliabilist can plausibly answer: the process types that have the dispositional property to produce true beliefs, used in epistemically safe conditions. This answer, of course, does not provide *one* relevant type-individuation and reference class, but it gives us some way to determine *a set* of relevant type-individuations and reference classes that meet the requirements of DRK. Whether one type-individuation is more or less reliable (in frequentist sense) than others does not matter as long as they all meet the necessary conditions for knowledge. Conee and Feldman (1998, 4–5) states three necessary conditions for a solution to the generality problem, (i) it has to be principled, (ii) it has to make epistemically defensible classifications, and (iii) it has to remain in the spirit of reliabilism. The answer DRK provides is principled, in that it even provides necessary and sufficient conditions for the type-individuation of the cognitive process (it is necessary that it has the dispositional property to produce true beliefs), and the reference class needs to be epistemically safe. It seems to be making epistemically defensible classifications as well, seeing as it avoids many of the counterexamples that standard reliabilism cannot account for. Lastly, is DRK in the spirit of reliabilism? I think so, but the argument for this will have to wait until the penultimate section of this paper.¹¹

It is unclear whether this approach is ultimately attractive, since one would still have to say something about how we determine which process types have the dispositional properties to produce true beliefs, which seems far from easy to answer (the best case scenario would be that it is an empirical question, the worst case scenario would be that it is an impossible-to-answer question, in that we cannot ultimately, perhaps, verify the truth of our beliefs with reference to some absolutely certain epistemic process). As with every other problem discussed here, the main aim is to give a plausible initial response. I hope to at least have given some credence to the idea that DRK *could* provide a solution to the generality problem.

¹¹See also Becker (2008) for a similar approach to the generality problem.

2.6. Lottery cases

Lottery-related problems arise for the classical conception of reliabilism that take reliability to be about producing a positive truth ratio. In other words, reliable processes are ones that produce significantly more true beliefs than false beliefs. The problem is that in a lottery where one ticket is winning out of, let's say, 1000 tickets total, then one can reliably form the belief that the ticket one draws is a losing ticket. Over time, this process would produce more true beliefs than false beliefs, and so the frequentist-reliabilist would have to concede that this is a way of knowing. But at the same time, we can tell that we have no knowledge whatsoever about whether the ticket drawn is lost. A principle like, "The obvious moral is that one is *never* warranted in asserting a proposition by its probability (short of 1) alone" (Williamson 2000, section 11.3) seems to fit this intuition. The problem in the lottery case is that we seem to permit jumps from probability statements to statements of fact. Knowledge is factive, and it implies the truth of its related proposition, and no probability short of 1 can allow for this entailment. Moreover, if one is permitted to make this jump, one is able to derive the same knowledge about every ticket in the lottery, which will inevitably lead to a false belief. It is also odd, then, to view an epistemic process as legitimate if it by necessity leads to falsity.

The problem takes a more sophisticated form in Adler's version of the lottery problem (2005, 446–447):

A company that manufactures widgets knows that exactly one out of every thousand of their products suffers a singular defect as a by-product of (an ineliminable imperfection in) their excellent – and much better than average – manufacturing system. Whenever there is a defect, it is sufficiently glaring so that customers recognize it and complain. Some managers would like to reduce the percentage of defects. The plan is to introduce a special detector, well designed to read 'OK' just in case the widget is not defective. The detector is to be applied subsequent to the normal manufacturing process to locate any defect before the widgets are shipped to the stores.

Smith and Jones, who both know of the one in a thousand defects, are each given a detector. Batches of widgets are randomly sent to one and later, without either one's knowledge, to the other. [...]

As each widget comes to his station, Smith momentarily glances away from the video game he is playing to stamp it 'OK', expressive of his corresponding (degree of) belief, while wholly ignoring his detector. As Smith knows, out of each batch of 1000, he is guaranteed to be correct 999 times by this method (and so better than by use of the detector, as explained next).

Less so Jones. Knowing of her manager's well founded confidence in the detector, Jones applies it to each widget carefully and skillfully, and assigns it 'OK' (mostly) or 'Defect' (rarely) according to her determination. Given the complexity of operating the detector and normal human limits, the probability of an error in any evaluation is .003, though Jones is a first-rate technician. [...]

Smith's manager regularly finds him playing video games, and threatens his job. Smith protests that his method yields a truth-ratio that is very high and better (and faster) than that of the esteemed Jones. Smith may add a zinger to his protest: for each widget where his and Jones' judgements differ, he would be willing to bet the manager (at even odds) that he is correct. (The laborious process noted above will settle bets.) Still, the manager remains unmoved.

Now, if epistemic evaluations were about reliability alone, it is unclear why Smith is being irresponsible, but not Jones. We can ask more generally, is the indignation of Smith's manager warranted? It seems so! There is a sense in which Smith, despite using a process that is more frequentist-reliable than the process of Jones, fails to act responsibly and so also fails to fulfill his work duties. The relevant argument would run something like this:

- (1) $K = R$
- (2) If one acts epistemically irresponsibly, one cannot have knowledge.
- (3) Smith acts epistemically irresponsibly.
- (4) Smith's method is reliable.
- (5) Smith has knowledge and does not have knowledge (via substitution of 4 via 1, and modus tollens using 2 and 3).

While Smith's job is in part to use a process that makes it very improbable that his widgets are defective, he has also been tasked to use his epistemic abilities to ensure that the widgets are not defective. He may do the former better than Jones, but he completely fails to do the latter. The problem with standard reliabilism is that it fails to account for this intuition. Standard reliabilism cannot give the reading that a person is epistemically blameworthy as long as their process – no matter how ridiculously irresponsible it seems to be – produces a high ratio of true beliefs. DRJ can at least provide some account for the idea that Smith has done something to warrant potentially losing his job insofar as he is not using a cognitive process type disposed to produce true beliefs. And so, neither knowledge nor justification is obtained. From Smith's manager's point of view, Smith can be seen as not using a cognitive process type disposed to produce true beliefs because, again, he is not doing anything to acquire true beliefs *specifically* about the widget in front of him. He is not using the cognitive powers available to him (indirectly via the use of the instrument that checks whether the widget is defective). Just as in the case with clairvoyance (unless stipulated to be in virtue of natural forces), we can now say something about why it is that Smith is not epistemically justified in virtue of his epistemically irresponsible behavior. In both cases, the epistemic agents (Norman and Smith) fail to use processes endowed with the dispositional property of producing true beliefs. If part of your job is to do just that, then it is highly irresponsible to deliberately avoid it, even if your results output a higher relative frequency of the ultimate desired outcome (to not sending out defective widgets).¹² Let us illustrate this point by inputting DRJ into the argument above. Now we fail to engender a contradiction:

- (1') Justification = using a CPT in safe conditions.
- (2') If one acts epistemically irresponsibly, one cannot be justified (and therefore cannot be in a state of knowledge).
- (3') If Smith does not use a CPT, then Smith acts epistemically irresponsibly.
- (4') Smith does not use a CPT (and so Smith is not justified).
- (5') Therefore, Smith does not have knowledge (Via modus ponens, 4', 2'–3').

¹²Adler (2005, 448) tells us that the lottery problem is about evoking comparative judgments, and the reliabilist needs to tell us why Jones is epistemically better off than Smith. Since Smith does not even form beliefs about individual widgets, this would be one simple way of accounting for why Jones is epistemically better off. If he were to form beliefs about each widget, he would know that the beliefs acquired, taken jointly, are inconsistent. If he accepts that result, he would be irrational, and so could not be ascribed with a CPT with the disposition to produce true beliefs.

For the lottery case, the same point can be made. In general, picking a lottery ticket is in no way a cognitive process type disposed toward true beliefs. Given a lottery where 500 out of 1000 tickets are winning tickets, one would not be able to make the claim that one had some privileged path to acquiring a true belief. Lotteries are not causally connected to the formation of true beliefs in the right kind of way to count as a cognitive process type with the property to produce true beliefs. The external circumstances could make a process of picking a lottery ticket truth-conducive, if one also knows the ratio of winning to losing tickets, but this would not give one the propensity to acquire true beliefs regarding lotteries in a more general sense. The point is this: A cognitive process type disposed to produce true beliefs can be used to produce true beliefs in a way that guessing the outcome of lotteries cannot. Some lotteries may put you in rather epistemically safe conditions such that it is going to be very likely that your belief that you picked a losing ticket is going to be true. But if a lottery makes it so that one has a .5 chance of losing, then it becomes clear why it is no longer an epistemic process; you are clearly not disposed to produce true beliefs in this case, because your cognitive process is simply not truth-conducive (in the dispositional sense). Safety can do a lot, but it does not account for one's belief being justified. So even if one outputs a high ratio of true beliefs in some forms of lotteries by believing that one has a losing ticket every time one buys a ticket, it does not count as a state of knowledge because it is not the right kind of cognitive process type, because it lacks the dispositional property to produce true beliefs.

Making an epistemic environment incredibly epistemically safe is one way to ensure that true beliefs are acquired without requiring anything of the epistemic agent. Letting a person know the odds of a lottery would be one way to do this. Another way would be this: Consider Mary, she has noticed a strange pattern. Anytime she guesses that something is the case, it turns out to be true. She wakes up in the morning, she guesses that it is raining outside. She gets up and looks out the window and sees that it is in fact raining. She guesses that her husband made her breakfast, and it turns out that he did, and so on. Mary keeps noticing these types of events and realizes that there is a pattern such that anytime she makes uninformed guesses about ordinary world events, they turn out to be true. Because of this, eventually she starts to believe whatever she guesses. Now, consider further the idea that unbeknownst to her, a rather silly demon is in charge of making each of her guesses true in order for her to start believing that her guessing is a reliable indicator of the way the world is. Mary's "knowledge" is epistemically safe. Nonetheless, she does not do what is required of her, which is to say, she does not use a type of process with the dispositional property to produce true beliefs. The reliability of her guesses is merely a form of dispositional mimicry. We know that she does not use a CPT with the dispositional property to produce true beliefs, and so she is not justified in her guess-related beliefs, and therefore, she also does not have knowledge. Similarly, one cannot acquire knowledge via lotteries since one is being cognitively irresponsible (even if one is in highly epistemically safe conditions). This irresponsibility can be explained by the fact that one's supposedly epistemic process lacks the right kind of dispositional property. Games of chance, no matter how favorable, cannot count as knowledge, for this reason.¹³

¹³As an alternative response, Pritchard (2005, 128 & 2012, 254) shows rather convincingly that safety can rule out lotteries as cases of knowledge, since in lotteries, picking the winning ticket may be very unlikely, but it is nonetheless a very near-by possible world, and so one cannot count picking lottery tickets as a safe way to form a belief.

3. The spirit of reliabilism – a dispositional approach

I have here intended to give preliminary responses to a wide range of “counterexamples” and explanatory problems facing classical reliabilism. While some of the problems are handled with ease and somewhat conclusively, others require more effort and are less conclusive. I nonetheless hope that the plausibility of DRK and DRJ has been increased by the fact that a number of plausible answers could be given to many of the biggest problems facing reliabilism in the literature to date.

The question remains, though, whether a dispositional and safety conception of knowledge is reliabilist in spirit. As Bach (1985) notes, Armstrong’s (1973) thermometer view of noninferential knowledge is too strong. Reliabilism cannot maintain a probabilistic relationship between justification and truth if justification consists in a “law-like connection” (ibid., 166). Reliability, on such a view, would entail truth. But it is unclear whether there are types of processes in this world that, whenever they are triggered, entail a true belief. Necessary connections of this kind are perhaps not to be found in the world at all. To demand such a connection in order for one to have knowledge seems too strong. Such a demand would essentially be Cartesian, and our claim to knowledge would be threatened altogether, or at the very least highly constrain the types of states we could count as states of knowledge. To lower the demand, then, we can instead view the connection between reliable processes and true beliefs as probabilistic. This relatively lower demand coheres well with C. I. Lewis’, who writes (1956, 369): “If the reality of knowledge required that [. . .] every given appearance was an index to some uniformity which could be predicated with certainty, then it would not be plausible that there is any such thing as knowledge.” We need only to modify the language by Lewis to get the same point for reliabilism. If every use of a reliable process was an index to a true belief that could be predicated with certainty, it would not be plausible that there is such a thing as knowledge. To avoid this, we must maintain the spirit of Goldman’s (1979) and Swain’s (1981) conceptions of reliability, where the latter is understood in terms of probabilistic relationships between certain types of cognitive processes and the acquisition of true beliefs.

But how are we to understand objective probabilities? Hájek (2023) brings up two well-known ways in which to understand probabilities (Goldman and Beddor 2021, also mentions these as being the two ways in which to interpret “reliability”).¹⁴ One way is to view probabilities as relative frequencies of events. On a frequentist understanding, if one wants to figure out whether a process is truth-conducive, one simply observes how many times the process actually produces a true belief, then divide this number by the total number of times the process was used. Another way to understand probabilities is to view them as dispositional properties of objects in the world (as propensities of the world itself). Hájek (ibid.) notes two different ways in which to understand the idea that there are propensities of real-world objects underlying observed relative frequencies. One comes from Peirce, who says that propensities are dispositional properties of individual objects. The other comes from Popper, who says that propensities are the properties of entire chance set-ups. When the propensity comes about through the properties of all the objects in an entire chance set-up, it means we attribute the disposition not merely to one object, but all of the objects relevant to the chance set-up.

Dispositional-reliabilist knowledge involves both kinds of propensities. On the one hand, in order to have a justified belief, one needs to use the cognitive processes with the right kind of propensities (the propensities to produce true beliefs). In order to have

¹⁴Goldman (2021, section 1.1) himself writes that reliability can be understood in terms of either frequencies or propensities. Another notable early adopter of a propensity view on reliability can be found in Schmitt (1983).

knowledge, one must be in the right kind of chance set-up, meaning that one has to be in safe epistemic conditions in order to actually yield a high ratio of true beliefs. Using a CPT with a dispositional property (I. above) captures the Peircean conception of propensities, whereas safety (II. above) captures the Popperian idea of propensities. When one uses a process with an intrinsic dispositional property to produce a belief, then that belief is justified. But it is only when the entire chance set-up is truth-conducive that one has knowledge. If safety is necessary for the entire chance set-up to be truth-conducive, and knowledge requires truth-conduciveness (as reliabilism states), then safety becomes necessary for knowledge.

This is why DRK and DRJ remain in the reliabilist spirit. Knowledge is probabilistic, as Lewis was among the first to argue, and as Goldman and Swain (separately) later systematized. What does it mean for something to be probabilistic? The interpretation chosen here: probabilities pertain to dispositions in our world in two ways, (1) individual dispositions of objects and (2) dispositions of systems of objects. These two ways of understanding probabilities, applied to reliable processes, cohere with DRJ and DRK, respectively, therefore dispositional reliabilism maintains the spirit of reliabilism. In actuality, the propensity understanding of reliability may be one of few ways in which to improve our understanding of what it means for a process to be reliable.¹⁵ In an ideal world, perfectly reliable indicators analyzable in terms of conditional analyses would be the sources of our knowledge, but it is not clear whether there are perfectly reliable indicators of this sort in the world. When one acquires knowledge, we could say that it is in virtue of the success of such reliable indicators. But we should not demand that the reliable indicator be an infallible tracker of truth before the user of this indicator be regarded as holding justified beliefs. When we move from reliable indicators to reliable processes, I believe we more accurately capture the way in which knowledge is actually acquired: by way of fallible – but nonetheless very reliable – processes.

4. Concluding remarks

I have attempted to give a new analysis of reliabilist knowledge that aimed at a more specific understanding of the notion of probability used in “reliable”. Dispositions and safety together fill this explanatory (conceptual) gap. By filling this gap, many problems for reliabilism now look to have a clear solution. In one fell swoop, if this account of knowledge were to miraculously withstand further scrutiny, reliabilism would gain much-needed conceptual underpinnings capable of solving problems that have irked the reliabilist for decades.

Acknowledgements. I want to express my gratitude to the people who read, commented on, and/or discussed various parts of the manuscript: Erik J. Olsson, Robin Stenwall, Andreas Stephens, and Sarah Köglspurger. Your insights and helpful suggestions improved both the content and style of the paper greatly. I also want to thank the anonymous reviewers for their astute criticism, which helped me avoid making a couple of substantial errors.

¹⁵Popper (1990, 12) is incredibly instructive when it comes to our understanding of probabilities. Statistical averages, he says, remain stable over time, and something needs to explain this fact. The notion that there are objective propensities in the world does just that. If this is true, is not the frequentist conception of reliability not a mere placeholder for some physical propensity? If so, should one not spell the physical propensity out as the very thing that underlies the idea that knowledge consists, essentially, in a special kind of probabilistic relationship between the acquisition of true beliefs and the use of reliable belief-forming processes? I believe the answer to both questions should be in the affirmative.

References

- Adler J.E. (2005). 'Reliabilist Justification (or Knowledge) as a Good Truth-Ratio.' *Pacific Philosophical Quarterly* 86(4), 445–58.
- Alston W.P. (1993). 'Epistemic Desiderata.' *Philosophy and Phenomenological Research* 53(3), 527–51.
- Armstrong D.M. (1973). *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.
- Bach K. (1985). 'A Rationale for Reliabilism.' *The Monist* 68(2), 246–63.
- Becker K. (2006). 'Reliabilism and Safety.' *Metaphilosophy* 37(5), 691–704.
- Becker K. (2008). 'Epistemic Luck and the Generality Problem.' *Philosophical Studies* 139, 353–66.
- Bird A. (1998). 'Dispositions and Antidotes.' *The Philosophical Quarterly* 48(191), 227–34.
- Bird A. (2020). 'Possibility and the Analysis of Dispositions.' *Philosophical Inquiries* 8(1), 83–96.
- BonJour L. (1980). 'Externalist Theories of Empirical Knowledge.' *Midwest Studies in Philosophy* 5(1), 53–74.
- Brandom R.B. (1998). 'Insights and Blindspots of Reliabilism.' *The Monist* 81(3), 371–92.
- Cohen S. (1984). 'Justification and Truth.' *Philosophical Studies* 46, 279–95.
- Cohen S. and Lehrer K. (1983). 'Justification, Truth, and Knowledge.' *Synthese* 55(2), 191–207.
- Comesaña J. (2002). 'The Diagonal and the Demon.' *Philosophical Studies* 110, 249–66.
- Comesaña J. (2006). 'A Well-Founded Solution to the Generality Problem.' *Philosophical Studies* 129(1), 27–47.
- Comesaña J. (2009). 'What Lottery Problem for Reliabilism?' *Pacific Philosophical Quarterly* 90(1), 1–20.
- Conee E. and Feldman R. (1998) 'The Generality Problem for Reliabilism.' *Philosophical Studies* 89, 1–29.
- Dretske F.I. (1970). 'Epistemic Operators.' *Journal of Philosophy* 67(24), 1007–23.
- Gettier E. (1963). 'Is Knowledge Justified True Belief?' *Analysis* 23(6), 121–23.
- Goldman A. and Beddor B. (2021). 'Reliabilist Epistemology.' In E.N. Zalta (ed), *The Stanford Encyclopedia of Philosophy*. Stanford: Stanford University.
- Goldman A. I. (1986). *Epistemology and Cognition*. Harvard University Press.
- Goldman A.I. (1976). 'Discrimination and Perceptual Knowledge.' *Journal of Philosophy* 73(20), 771–91.
- Goldman A.I. (1988). 'Strong and Weak Justification.' *Philosophical Perspectives* 2, 51–69.
- Goldman A.I. (1979). 'What Is Justified Belief?' In G.S. Pappas (ed), *Justification and Knowledge*. Dordrecht: D. Reidel.
- Greco J. (1999). 'Agent Reliabilism.' *Philosophical Perspectives* 13, 273–96.
- Greco J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Gundersen L. (2010). 'Tracking, Epistemic Dispositions and Conditional Analysis.' *Erkenntnis* 72(3), 353–64.
- Gundersen L.B. (2003). *Dispositional Theories of Knowledge : A Defence of Aetiological Foundationalism*. Aldershot: Ashgate.
- Gundersen L.B. (2012). 'Knowledge, Cognitive Dispositions and Conditionals.' In *The Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press.
- Hájek A. (2023) 'Interpretations of Probability.' In E.N. Zalta and U. Nodelman (eds), *The Stanford Encyclopedia of Philosophy*. Stanford: Stanford University.
- Handfield T. and Bird A. (2008). 'Dispositions, Rules, and Finks.' *Philosophical Studies* 140, 285–98.
- Kvanvig J.L. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
- Kvanvig J.L. (2010). *The Swamping Problem Redux: Pith and Gist*. Oxford: Oxford University Press.
- Lehrer K. (1990). *Theory of Knowledge*. London: Routledge.
- Lewis C.I. (1956). *Mind and the World-Order: Outline of a Theory of Knowledge*. New York: Courier Corporation.
- Merleau-Ponty M. (2014). *Phenomenology of Perception*. London: Routledge.
- Mumford S. (1998). *Dispositions*. Oxford: Oxford University Press.
- Mumford S. (2009). 'Ascribing Dispositions.' In *Debating Dispositions: Issues in Metaphysics, Epistemology and Philosophy of Mind*, pp. 168–85. Berlin: De Gruyter.
- Popper K. (1990). *A World of Propensities*. Bristol: Thoemmes.
- Pritchard D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard D. (2012). 'Anti-Luck Virtue Epistemology.' *The Journal of Philosophy* 109, 247–79.
- Schmitt F.F. (1983). 'Knowledge, Justification, and Reliability.' *Synthese* 55, 209–29.
- Sosa E. (1992). 'Generic Reliabilism and Virtue Epistemology.' *Philosophical Issues* 2, 79–92.

- Sosa E.** (1999). 'How to Defeat Opposition to Moore.' *Philosophical Perspectives* **13**, 141–53.
- Swain M.** (1981). 'Justification and Reliable Belief.' *Philosophical Studies* **40**(3), 389–407.
- Unger P.** (1968). 'An Analysis of Factual Knowledge.' *The Journal of Philosophy* **65**, 157–70.
- Williamson T.** (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Balder Edmund Ask Zaar holds a Master's degree in Theoretical Philosophy from Lund University, where he is also currently a Ph.D. student. His academic interests include epistemology in general and reliabilism and its problems in particular.