DISPOSITIONAL RELIABILISM AND ITS MERITS

Balder Edmund Ask ZAAR

ABSTRACT: In this article I discuss two counterexamples (the New Evil Demon Problem and Norman's Clairvoyance) to reliabilism and a potential solution: dispositional reliabilism. The latter is a recent addition to the many already-existing varieties of reliabilism and faces some serious problems of its own. I argue here that these problems are surmountable. The resulting central argument of the article aims to demonstrate how viewing reliabilism as an intrinsic dispositional property solves many of the issues facing reliabilism to date.

KEYWORDS: dispositional reliabilism, counterexamples to reliabilism,
new evil demon problem, accidental reliability

1. Introduction

Reliabilism can easily be said to capture what is essential about epistemic justification. It does so by emphasizing that justification is about a certain kind of reliable relation to truth such that if one is justified in holding a belief that P, then P is also very likely to be true. Were one to make a stronger claim, for instance by saying that the relation between epistemic justification and truth is such that if one has a belief that P, then P is invariably true, then one is also incapable of handling the very plausible scenario wherein one is justified in holding a belief that P without P being true. Reliabilism effectively solves this problem by offering the next best thing: epistemic justification is when one's true belief is reliably (but not infallibly) produced. This still leaves open the possibility that one's belief is produced with a reliable method, yet that belief is false. Reliabilism thus offers the closest possible relation that a justified belief may be said to hold to truth without thereby making the relation logical or nomological. Reliability alone, however, has also been shown to be neither necessary nor sufficient to account for justification, and so the trouble begins.

This article amounts essentially to a defense of reliabilism by defending 'dispositional reliabilism.' I will begin by introducing in a bit more detail what I call 'standard reliabilism' and two of its more prominent counterexamples in order to then turn to a longer discussion of dispositional reliabilism and attempt to show how the theory is capable of facing the purported counterexamples head on.

The standard reliabilist theory of justification states that a belief is justified if and only if it has been formed by way of a reliable process. Along the same line, reliabilist theories of *knowledge* necessitate that a true belief is the result of a reliable process if it is to count as a state of knowledge (Goldman 2021). For a belief-forming process to be reliable means that it is truth-conducive. That a process is truth-conducive, in turn, means that the process has to have a high probability of producing true beliefs or, put differently, it has to have a high truth ratio; the process has to produce a higher ratio of true beliefs compared to false beliefs. For many, this amounts to the essence of whatever it is that takes us from mere true belief to genuine knowledge. That is to say, the reliabilist wants us to think that regardless of how one conceives of epistemic justification, it is essential that states of justification are truth-conducive, and thus reliable, or else we would have no reason to view having a justified true belief as more valuable than having a mere true belief.

There are two types of direct counterexamples to the reliabilist conception of justification.[1] One of them is the so-called New Evil Demon (NED) Problem, which first appeared in Lehrer and Cohen (1983) and Cohen (1984). The counterexample hinges on what I call the 'internalist intuition.' The intuition arises when we consider a world just like ours except for the fact that there is an evil demon that ensures that all the normally reliable processes only engender false beliefs. Their belief-acquiring processes, in other words, are no longer reliable. Yet, the internalist would say, the inhabitants of the demon world are nonetheless justified in holding their beliefs insofar as they are doing things such as appealing to the best available evidence, and so they are still being maximally epistemically responsible, and are, in a sense, still justified in holding various beliefs about their world. What counts as being out of the control of the NED-worlders simply cannot be used to undermine their status of being justified. Now, if the NED-worlders are as justified as we are, then the following reductio argument can be constructed to undermine reliabilism:

1. The NED-world inhabitants cannot acquire beliefs reliably (NED-world Stipulation).
2. A belief is justified if and only if it has been formed by way of a reliable process. (Reliabilist Assumption)
3. The NED-world inhabitants' perceptual beliefs are as justified as our own. (Internalist Intuition)

[1] To my knowledge at the time of writing this.

4. Therefore, the perceptual beliefs[2] of the NED-world inhabitants have been produced by reliable processes. (1, 4; ⊥)

The other counterexample is arrived at through considering worlds where a type of reliable process is merely reliable for seemingly accidental reasons. Consider the following quote by 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.

Now what kind of trouble does this cause for the reliabilist? It shows that accidentally reliable processes, the ones that are also highly irresponsible processes of belief-formation, may nonetheless amount to states of being justified simply in virtue of being truth-conducive processes. One may thus use reliable processes to acquire a belief that P without thereby having good reasons to believe that P. This is far from ideal.

Both counterexamples indicate that reliabilism alone is neither necessary nor sufficient to account for whatever it is that brings us from true belief to knowledge. The NED-problem show us that reliability is not necessary in order for a person to be justified. The problem of mysterious or non-normal reliabilism shows us that reliabilism is not sufficient to bring us to a state of justification or knowledge. As long as there is something accidental or seemingly irresponsible about one's reliably formed belief, that belief cannot be seen as justified. Facing such powerful counterexamples, one would not be amiss to think the reliabilist project to be rather hopeless.

In the proceeding article I am going to attempt to show how adopting dispositional reliabilism undercuts both types of counterexamples and as a result preserves a form of standard reliabilism, albeit a more specified version of it. Let us now make clear what dispositional reliabilism is and how it purports to fix the problems of standard reliabilism.

[2] Perceptual beliefs are normally taken to be justified, which is why I use them here, but any other kind of belief that we paradigmatically take to be justified or reliably acquired could be used (so instead of a perceptual belief, it could be one arrived at through using sound reasoning, and so on).

2. Why Consider Dispositional Reliabilism?

Dispositional reliabilism arises out of a rather plausible analogy argument, which will be presented shortly. But let us first introduce the general notion and how it serves as an improvement on standard reliabilism and other prominent varieties of reliabilism.

The idea, in short, is to view reliable processes as possessing particular kinds of dispositional properties. Dispositional reliabilism then posits that to use a reliable process is to use a process *disposed* to produce a high ratio of true beliefs. So, for example, to be engaged in a reliable perceptual process resulting in a true belief is to use a process that has the dispositional property to produce a high ratio of true beliefs. This minor modification of standard reliabilism means, according to Baysan (2017), that we need not 'weaken, relativize, or indexicalize' (paraphrasing Baysan 2017, 42) standard reliabilism in order to solve the NED-problem (as well as the problems surrounding how to view accidental reliability's relation to being justified). The kind of relativization and weakening here involves views such as 'home-world' or 'actual-world reliabilism' (Majors & Sawyer 2005), indexical reliabilism (Comesaña 2002) and normal-world reliabilism (Goldman 1986, 107). Under the category of 'weakened' versions of reliabilism I would also include two-concepts responses, such as Goldman's strong and weak justification (1988) and Sosa's (2003) apt and adroit justification. A single-concept response would be preferable simply via considering something like Grice's razor, but the two-concepts responses also fail in their own right. Goldman's two-concepts response involves something like the following (this exact formulation can be found in Majors & Sawyer 2005, 270):

> **(S/W)** Strong/Weak Justification: Justification consists either in reliability in the world the subject happens to inhabit (strong justification), or in unreliable but cognitively responsible belief (weak justification).

While this view accounts for the internalist intuition, it can be said to fail to satisfy the crucial desideratum of externalist epistemology. Perhaps by calling the form of justification that the internalist intuition appeals to 'weak,' Goldman still captures the fact that truth-conduciveness in the world a subject happens to inhabit is what epistemic justification primarily aims for. But in allowing cognitively responsible beliefs (which can exist without any relation to the truth whatsoever) to count as justified beliefs, we seem to have conceded too much. Standing in a particular relation to the way things are is no longer essential to our notion of epistemic justification. We thus end up with two concepts of justification. One (the strong) which leads to a contradiction (see the NED-argument above), another which, to an externalist, is no theory of justification at all, since complete cognitive responsibility can be in place (as in the NED-world) without an epistemic subject having any sort of relation to the way things are. That is to say, weak justification lacks the requisite truth relation, and as such, is in an externalist framework, no theory of epistemic justification at all. While this kind of conciliatory response has many merits in its own right and might be acceptable given more extended consideration, I take this kind of approach to be, at best, a last resort. As long as we can remain within single-concept theories of justification, we should do precisely that.

With Sosa's two-concepts theory things are not looking any better. The case against it will be presented in brief (it has been convincingly undermined already, cf. Graham 2016 and Majors & Sawyer 2005). Sosa calls the two ways in which one can be justified apt and adroit justification. The former is the justification one has when using intellectual virtues to arrive at beliefs where using the virtues yields a high ratio of true beliefs in the world of usage; the latter is the kind of justification one has when using intellectual virtues that yield a high ratio of true beliefs in the actual world. Consider again the case of a user of clairvoyance in a world where clairvoyance happens to be reliable for accidental reasons. Does Sosa's theory account for this scenario? Adroit justification clearly does not work since clairvoyance is not reliable in the actual world (and actual world reliabilism generally cannot judge whether someone is justified in a non-actual world since they could never in fact be justified merely in virtue of being in the wrong world). Apt justification, on the other hand, is accounted for, but it seems to face the exact same issues that standard reliabilism faces. Accidentally or contingently reliable processes are just not what we typically take to be responsible ways to acquire beliefs. Moreover, a highly irresponsible way of forming a belief cannot be used to justify a belief. Actual-world reliabilism (either adroit or apt) is not able to account for this problem.

Now these types of responses have been brought up in part to show that plurality of concepts does not necessarily lead to a working theory of justification, but also to illustrate that if a single-concept response was indeed possible, it would be doubly preferable; not only in virtue of Grice's razor, but because these two-concepts responses seemingly do not work. Let us now turn to the question of the plausibility of dispositional reliabilism and see how it faces the problems that other varieties of reliabilism are seemingly unable to handle.

3. The Problems Facing Dispositional Reliabilism and How to Move Forward in Spite of Them

Why should we think that dispositional reliabilism is plausible? It is meant to serve as a solution to the NED-problem. But the NED-worlders, by stipulation, are not able to produce true beliefs when using their perceptual processes, so how can a process be disposed to produce true beliefs without being capable of producing beliefs? The underlying assumption supporting dispositional reliabilism is that one can be a bearer of a dispositional property without manifesting the property for entirely contingent reasons, even when these reasons are systematically present. This puts into question what reliabilism actually consists of, and how it arises. Is it an internal state of a person which is central or an internal state mixed with a particular environment? The latter is also the point of tension that Madison (2021) picks up on, which will drive the coming discussion. But let us first go through the analogy argument in favor of dispositional reliabilism to explain how it can overcome the immediate problem noted above regarding how one can be disposed to produce true beliefs but never doing so.

The argument runs roughly like this (ibid., Baysan 2017, 44-45):

(1) A vase can be fragile without ever breaking simply by never being struck (or going through any other event which could break it).

(2) Further, there could be vases which never break despite being struck, let us say if there is some kind of magic spell on it which prevents it from breaking when it otherwise would have.

(3) The vase is nonetheless fragile, since if it were not for the protective spell (a contingent fact about the vase), it would have manifested its fragility by breaking on being struck.

(4) Similarly, if (3) is a possible state of affairs then the following state of affairs is also possible: say *a* is a reliable belief-forming process; *a* is used by a subject *S*; *a* nonetheless fails to produce true beliefs for *S* and does so systematically.

If the vase is fragile without being breakable, then, similarly, our perceptual processes can be reliable without outputting true beliefs. If this is plausible, it could then solve the NED-problem by using standard reliabilism understood dispositionally. The NED-worlders can then truly be said to be using processes that have the dispositional property of being reliable with the caveat that they are being systematically prevented from manifesting this property by the evil demon. If epistemic justification consists in dispositional reliability, then it also explains why the NED-worlder is as justified as their actual-world counterpart: their perceptual processes share a certain kind of dispositional property. We also solve the issue of accidental reliability. The person using clairvoyance is in fact not using a reliable process insofar as the process being used does not have the dispositional property of being reliable. Its manifest reliability is entirely contingent. Alternatively, we could view the case of clairvoyant reliability as a case where a world has acquired clairvoyance-waves and receptor or similar strange phenomena that could indeed posit the existence of a non-contingent form of reliability in that world. But in that case, we move further and further away from the accidental nature of such a scenario, and we can no longer claim that such processes are irresponsible to use. Instead, we can now claim that the accidental nature of the reliability is based on the fact that the process lacks a vital dispositional property. So far so good.

But what are we to make of the idea that someone in the NED-world is using reliable processes? That is, how do we specify the notion of reliability so that a process can be said to be reliable without thereby producing a high ratio of true beliefs in certain scenarios? Madison (2021, 197-198) gives two suggestions.

One suggestion is to view reliability as a property ascribable to processes that have a track record of producing true beliefs. The frequency with which the process produces true beliefs compared to false beliefs must then meet some threshold in order to be seen as reliable. Now this does not work since the NED-problem, as Madison points out, ensures that the track record of the processes used by the NED-worlders is such that it has not produced a high ratio of true beliefs. The point is broader, also, in that we cannot take producing a high proportion of true beliefs as either necessary or sufficient for reliability. It is conceivable that a process exists that is reliable, were it to be used, but that nonetheless is never used, therefore having used the process cannot be necessary for ascribing the property of being reliable. Along the same line, Madison (2021, 198) points out that one can use processes that are accidentally or 'luckily' truth-conducive – i.e., a student may guess all the answers on a test, through sheer luck be correct in those guesses, and be deemed to use a reliable process (although this is of course based on the type-process one is considering – guessing is not exactly a process that is reliable in general). These arguments may not be impossible to respond to, but together they at least make the plausibility of a functioning frequentist conception of reliability more problematic than its alternative.

These arguments suggest that reliability is better viewed in modal terms. It does not have to be the case that a process has already been shown to produce a high ratio of true beliefs (compared to false beliefs), but it is enough that the process *would* produce a high ratio of true beliefs if it were used.[3] A modal

[3] Goldman (1976, 771) himself, importantly, early on saw the importance of the counterfactual

conception of reliability, then, states that were one to use a reliable process, then one would yield a high ratio of true to false beliefs. Of course, if there is a posited demon which invariably blocks any perceptual belief from being true, a standard modal conception of reliability cannot work either. Since the NED-worlders are justified, that would mean they are using processes that, if they were used, would yield a high proportion of true beliefs (compared to false beliefs). But, seeing as it is stipulated in the NED-problem that such a favorable ratio of true to false beliefs cannot arise, this amounts to a clear contradiction. Although, with further specification, it is going to be argued here that modal reliabilism, interpreted through a realist dispositional framework with focus on intrinsic dispositional properties, is the best path forward. But let us first continue with Madinson's argument.

It is indeed impossible for our NED-world counterparts to yield true beliefs from their supposedly reliable processes. So what Madison, I believe correctly, points out is that Baysan fails to consider whether the dispositional properties under consideration are extrinsic or intrinsic properties (or a mix of the two). For the vase, the fragility can be said to be an *intrinsic* property of the vase; fragility is a property which comes from the vase's microstructure (as Madison puts it, ibid., 200).[4] The relational aspects of the vase as it is put under a protective spell, however, make it non-fragile in virtue of its *extrinsic* properties. The assumption Madison operates under here is that dispositional properties stem from the intrinsic

side of reliabilism: "a cognitive mechanism or process is reliable if it not only produces true beliefs in actual situations, but would produce true beliefs, or at least inhibit false beliefs, in relevant counterfactual situations. The theory of knowledge I envisage, then, would contain an important counterfactual component."

[4] Cf. something like Armstrong's (1993, 87-90) argument against the phenomenalist conception of dispositions. He reaches the conclusion that dispositions need support or arise from non-dispositional states. The argument for this is roughly that if a disposition is not due to intrinsic properties of an entity with a certain disposition, in that from the behaviorist/phenomenalist point of view one does not accept unobservables, then the disposition must be explained by its extrinsic properties (or by contingent connections between 'categorical properties and dispositional properties'). The extrinsic properties, however, are not observable either, which forces the phenomenalist to reject dispositions altogether – a position Armstrong takes to be too extreme. Thus, if we want to accept that things have dispositions, we also have to be realists about them. Being a realist about dispositions, in turn, seems to lead to a view where a thing's dispositional property is determined by its intrinsic properties. As Armstrong puts it (ibid., 88): "Dispositions are seen to be states that actually stand behind their manifestations". It is not an uncontroversial account, see for instance Mellor (1974, 164-5), but it nonetheless is at least plausible that dispositions in an object arise from that object's non-dispositional properties.

properties of a given thing, and that the intrinsic dispositional properties of our cognitive processes are not enough to yield genuine reliability.

Taking this distinction into account, we begin to see a disanalogy between the vase case and the case with the cognitive faculties – or so Madison claims. That is, supposedly two vases (one under a protective spell, one not) can share intrinsic properties and simultaneously have different conditions for manifesting their dispositions depending on the extrinsic properties of each vase. For the case of cognitive processes, Madison writes (ibid.):

> If two subjects are exact intrinsic duplicates, and have the same belief forming processes, these processes need not be equally reliable – for instance, the subjects might be in radically different environments, as the NED cases make vivid. Whether a process produces true beliefs is partly determined relationally. This means that whether a process is reliable necessarily depends on the environment in which it is used. Baysan seems to implicitly recognize this, as reliable belief-forming processes are described as tending to produce true beliefs, in the right circumstances.

What Madison stresses here is that reliability is not a wholly intrinsic property, and so the relationship between the vase example and the cognitive faculties example is not analogous, and so the argument Baysan posited fails. The result of this is that Baysan seemingly has to argue for a kind of *Modal Reliabilism*, akin to the varieties of reliabilism discussed in section 2 that are basically modified versions of standard reliabilism, relativized to special kinds of possible worlds. As Madison writes (ibid., 201): "In short, reliability is determined not only by the relevant belief forming process, but also the relevant environment, and the Dispositionalist response to the NED problem overlooks this." And so, the reliabilist has to type-individuate environments, as well as processes, in order to have genuine reliability (or in order to manifest a high ratio of true beliefs as acquired on the basis of using a reliable process).

But here we can question a few of the assumptions made by Madison.[5] First, let us consider the idea that the reliabilist has to type-individuate environments in order for reliabilism to be plausible. On this point there is a distinction to

[5] Another approach to Madison's argument which aims to uphold the analogy between vases and reliability would be to argue that fragility is not a property ascribable only on the basis of the intrinsic properties of an object. In a possible world where ceramic is the hardest material, for example, it seems that it would not be deemed to be fragile in the same sense a cognitive process would not be deemed reliable if there was an evil demon influencing things. Fragility seems to also be dependent on extrinsic factors such as forces and objects that are capable of instantiating the breaking of object disposed to breaking. Therefore, the analogy holds by viewing fragility as a mix of intrinsic and extrinsic factors as well.

highlight, which is the notion of reliability as compared to the notion of having a fully specified relation between a cognitive faculty and the environment it attempts to model, predict, understand, perceive, etc. Part of the virtue of a reliabilist notion of justification is that it is it not an infallibilist conception of justification. The relationship between justification and truth need not be one-to-one. All we have to do in order to be justified, according to the reliabilist, is to use cognitive faculties, processes, or methods, that yield a high ratio of true beliefs (in this case, using the modal account, the method *would* yield a high ratio of true beliefs if used). But the requirement that we have to specify the entire process to the point where there is no room for failure, that we have to guarantee that if one uses a certain process, then one is also going to yield a true belief, is not in the spirit of reliabilism as I understand it. What makes reliabilism attractive in part is that it is fallibilistic. That is, it is attractive because it does not require that we specify the full set of extrinsic properties that need to be in place in order to actually yield a true belief in each token use of the process. That a type of process yields a high ratio of true beliefs comes from the fact that the process allows, when the right circumstances are in place, one to gain knowledge of a state of affairs – not that it invariably does so in every conceivable situation. We are not always in favorable circumstances, and so there is no guarantee of gaining knowledge in each token use of a perceptual process (there are many ways in which to hinder a perceptual process from performing its proper function). Can this not be explained by the fact that reliability (of the modal variety) is an *intrinsic* dispositional property? The remainder of this section will attempt to provide reasons to answer this question in the affirmative.

Part of the problem with Madison's account is the assumption that dispositional reliabilism, in order to count as a form of justification, has to involve both intrinsic *and* extrinsic properties. Such a requirement seems to fly in the face of reliabilism insofar as it is normally taken to be a form of fallibilism. Reliabilism allows for error in the following sense: we might be in a situation where we use a type of process that normally produces true beliefs and acquire a false belief, but we need not worry about this as long as the process produces true beliefs in most case. If we have to guarantee, in any knowledge-seeking activity, that both relevant environmental/relational factors are present as well as making sure that the process involved has the desired dispositional property of being reliable, then it seems that we no longer have a notion of justification which allows for error. If, within the analysis of knowledge or justification, we systematically remove all environmental factors that may lead to error as well as use only processes that have a dispositional property of being reliable, then knowledge presumably cannot be fallible. Furthermore, one may not ever be able to be in a state wholly devoid of the possibility of doubt and therefore it is hard to see how there could be such a thing as knowledge instantiated in everyday epistemic situations. For a frequentist kind of reliabilism to arise, of course it is true that certain external conditions need to be in place in order for a process to produce mostly true belief. But perhaps what is important in order to be justified, is that one is engaging in the best possible processes one has at hand, i.e., those processes that are disposed to produce true beliefs, even if they end up never producing true beliefs, due to entirely extrinsic factors.

For if we on the other hand view justification as using processes that have the intrinsic dispositional property of being reliable, we seem to reach a theory of justification that explains the fallibility of knowledge as well as the internalist intuition. Reliability, I take it, does not have to be guaranteed by relational properties, it only has to contain the broader possibility of being reliable (by 'broader' here I mean that even if there are worlds in which some processes that are normally reliable are not reliable in those worlds, there nonetheless remains a possibility that they could turn reliable, were the relational properties that make the production of a high ratio of true beliefs impossible to disappear). Whether one has knowledge when one has a justified belief, then, could be said to be a contingent fact both in that it depends on whether the belief is true, but also on whether the intrinsic dispositional property of being reliable is in favorable conditions (that is, in conditions devoid of manifestation-blockers[6] such as evil demons or blindfolds). These external conditions, however, do not have to be taken as elements in the analysis of knowledge; they seem to come with the fact that epistemic agents inhabit mostly favorable epistemic conditions. So, while Madison is correct that extrinsic and intrinsic properties both need to be in place in order for there to be an observably truth-conducive process, this need not imply that reliabilism as a form of epistemic justification has to be understood as a mix of intrinsic and extrinsic factors. It may just as well be understood as consisting of primarily intrinsic factors; as properties of various cognitive processes or other methodological procedures (properties of various instruments used in experiments, and so on). If we view reliabilism in intrinsic and modal terms, it seems that Baysan's analogy nonetheless works.

There is also an added bonus with intrinsic or internal reliability in that one can analyze the dispositional property conditionally by including anti-masker and anti-mimic clauses (the latter being what happens in the clairvoyance scenario where a process is used but is not really disposed to produce true beliefs in virtue of

[6] Called a "masker" by Johnston (2012), "antidote" by Bird (1998).

the intrinsic properties of the process, but for seemingly entirely accidental reasons that mimic what a genuinely reliable process does). The NED-problem could then be viewed as a case where a disposition is masked by a demon's interference with the belief-acquiring processes, whereas the clairvoyance problem would then be a case where the disposition is mimicked (but where we can say that the dispositional property is not really there, and so cannot afford justification to a subject using such a process, thereby solving the problem of accidental reliability). To make the formulation of a conditional analysis of dispositional reliability a bit more precise: S is disposed to produce a high ratio of true beliefs when using process X if and only if S produces a high ratio of true beliefs given the use of X and there is no antidote or mimic present. This kind of understanding of reliability seems to preserve the fallibilism of standard reliabilism while undermining both purported counterexamples. It also avoids the criticism by Madison by taking the justification-conferring aspect of a reliable state to be intrinsic factors alone, thus plausibly upholding the analogy to Baysan's vase.

I will propose, then, that reliabilism is best viewed as an intrinsic dispositional property that a process has if it has the capacity to yield a high ratio of true beliefs in virtue of its intrinsic properties. The process, in some sense, has to be shown to be capable to provide a subject with information about the outside world in a way that is not accidental, in order to be viewed as a reliable process; there has to be proof of receptivity. So, we can say that even in the NED-world, the perceptual processes do have the capacity to yield a high ratio of true beliefs, although they cannot exercise this capacity due to the extrinsic properties that the world imposes on them. In virtue of their intrinsic properties, however, the processes are still reliable, but masked. We are now in a position to explain the internalist intuition with the help of an externalist framework, since we can say that the justified status of the NED-worlders is conferrable in virtue of the dispositional properties of their perceptual processes. Whether one is justified is still about factors that need not be present to the mind, and there is still an emphasis on the relation between justification and truth, only now focused more on the use of processes whose intrinsic properties are truth-conducive (in a modal sense).

So, to clarify, if we permit that justification consists in the intrinsic propositional property of being reliable, then the NED-problem is solved. For the same way the vase maintains its intrinsic fragility despite the sorcerer's protective spell, then, we could say that perceptual processes maintain their intrinsic reliability despite being in extraordinarily strange environments that mask their manifestation. On the flipside, we can also say that in the same way that the *fragile* vase is impossible to break, the reliable processes are incapable of yielding true beliefs (in that possible world). In both cases, these seemingly contradicting facts are merely due to the extrinsic factors involved, and so have no bearing on the dispositional status of the intrinsic properties. If this much can be permitted, we can now ascribe NED-worlders with intrinsic dispositional reliability without contradiction. Let us formulate intrinsic dispositional reliability (IDR):

> **IDR** Epistemic justification consists in using processes that are intrinsically disposed to yield a high ratio of true beliefs.

Let us now input this version into the NED-argument:

1. The NED-world inhabitant cannot acquire beliefs reliably. (NED-world Stipulation).
2. A belief is justified if and only if it was acquired via processes that are intrinsically disposed to yield a high ratio of true beliefs. (IDR)
3. The NED-world inhabitants' beliefs are as justified as our own (Internalist Intuition)
4. Therefore, the justified beliefs of the NED-world inhabitants have been acquired via processes that are intrinsically disposed to yield a high ratio of true beliefs. (3, 2)

Now there is no contradiction – being disposed to manifest a certain property does not mean one invariably does so (especially not in worlds that are epistemically unfavorable). We are also now in a position to better deal with the clairvoyance counterexample, in two separate senses. One problem with the possible world wherein clairvoyance is a reliable belief-forming process is that we would not view using such processes as being justified or responsible. Whereas it is a problem for standard reliabilism that there is reliability without justification, this is not a problem for IDR, since clairvoyance, as a state of mind, is not disposed towards yielding a high ratio of true beliefs in virtue of its intrinsic properties. So, even if clairvoyance happens to be reliable in this world, this cannot be in virtue of the dispositional property, and so we cannot say that the inhabitant of such a world is justified in using clairvoyance as a belief-forming process. With IDR we have a better idea of why it is that the clairvoyance reliability is accidental – it is only in virtue of unspecified extrinsic properties that Norman's clairvoyance is truth-conducive.

Alternatively, in the scenario suggested by Goldman (1988)[7] where the feeling of clairvoyance is coupled with new natural phenomena (clairvoyance

[7] To quote him directly (ibid., 62):

waves and clairvoyance wave receptors, for instance, which involves a real causal connection between the feeling and the phenomena in the world). Using clairvoyance in such a world would afford a subject the status of being justified in that the feeling of clairvoyance would indeed be disposed to produce true beliefs in virtue of its intrinsic properties. And so, it seems, the desideratum of having a non-accidental relation between justification and truth can also be maintained.

I would then make the case that intrinsic reliabilism is enough for justification. One is rarely in a position to verify the complete causal relationship between one's cognitive states and the environment; such a requirement would be too demanding. The conjecture here, then, is that in order to be justified it is enough to use a type of process whose intrinsic dispositional property allows the acquiring of true beliefs. While knowledge may consist in having a justified true belief, a justified belief need not be true, and this can be explained by having justification be tantamount to an intrinsic dispositional property of our cognitive faculties. The same way our cognitive faculties are disposed to produce a high ratio of true beliefs even if there is an evil demon systematically deceiving us, sugar is disposed to dissolve in water even if all water in a possible world is at an absolute zero.

The intrinsic dispositional property can now be said to be ascribable to the NED-worlders, yet nonetheless it cannot manifest itself due to the strange circumstances. If one were to remove the relational property of being 'influenced by an evil demon' for the NED-worlder, the intrinsic dispositional property would again manifest itself. In the same way, the vase would break if struck if the protective spell was removed. The analogy seems to hold, all that was needed was to heed the distinction between intrinsic and extrinsic properties (and perhaps reframe the NED-problem as a masking problem for a conditional analysis of

> Consider a possible non-normal world W, significantly different from ours. In W people commonly form beliefs by a process that has a very high truth-ratio in W, but would not have a high truth-ratio in normal worlds. Couldn't the beliefs formed by the process in W qualify as justified?
>
> To be concrete, let the process be that of forming beliefs in accord with feelings of clairvoyance. Such a process presumably does not have a high truth ratio in the actual world; nor would it have a high truth ratio in normal worlds. But suppose W contains clairvoyance waves, analogous to sound or light waves. By means of clairvoyance waves people in W accurately detect features of their environment just as we detect features of our environment by light and sound. Surely, the clairvoyance belief-forming processes of people in world W can yield justified beliefs.

dispositions and reframe the clairvoyance problem as a mimicking problem). While Madison's criticism was that genuine reliability required some kind of extrinsic property in order to guarantee reliability, the view espoused here is that a frequentist kind of reliability need not be guaranteed in order for one to be justified, instead one only needs to use processes that hold the dispositional property of being reliable, seeing as such a process would be what actually ends up making a frequentist conception of reliability possible. The cognitive faculties we have are reliable intrinsically, but not infallibly. They can even systematically be manipulated. Luckily, in normal conditions in our actual world, as far as we know, this is not the case, and so something like knowledge with all likelihood exists and we need not wade into skeptical waters.

Before concluding this article, I would like to add some comments in support of the analogy between the vase's fragility and our perceptual processes' reliability and formulate an argument in favor of dispositional reliabilism.

4. The Argument for Perpetual Dispositional Masking

In Mellor's In Defense of Dispositions (1978) there is relevant distinction between a thing being mortal and a thing being fragile. In the former case, there are implications regarding the future; the thing will die. In the latter, it is not necessary that the thing either has been broken or will break. Dispositions can thus be said to be different to other kinds of properties in a very clear way, i.e. (ibid., 159) in a way that also gives support to Baysan's analogy:

> [B]eing forty or mortal now has past or future consequences where being fragile or soluble does not. His past birth being what makes a man forty now, it must have him thirty ten years ago; similarly a man who is mortal now is bound to be mortal until he dies. We draw no such consequences from the present ascription of dispositions. A fragile glass may (or may not) be toughened by heat treatment at any time.

Moreover (ibid., 173):

> The safety precautions at our nuclear power station […] are intended to prevent an explosion by making impossible the conditions in which fuel would explode. It is ridiculous to say that their success robs the fuel of its explosive disposition and thus the precautions of their point.

This seems to add some support to the plausibility that one can have a dispositional property without it ever manifesting. If we accept that stimulus conditions for a certain disposition can be in place without the related disposition manifesting itself, we should also be able to accept that the reason for the failure of the disposition to manifest itself could be there in perpetuity. If we accept this, the following (similar to Baysan's) argument may be plausibly posited to sum up the discussion:

> (1) Having a disposition is compatible with it failing to manifest despite having met some stimulus condition (the situation may involve some kind of interference with the stimulus conditions, i.e., an evil demon).
>
> (2) There must be some reason as to why a disposition failed to manifest itself (Realist assumption about dispositions).
>
> (3) There is a possible world where the reason (the condition) for the disposition's failure to manifest itself is present in perpetuity.
>
> (4) Thus, there can be a possible world in which one can have a disposition despite its manifestation being impossible in that world.

While it would not be possible for NED-worlders to identify the underlying reliability of their perceptual states (which is stipulated anyway) due to them never manifesting their capacity to yield a high ratio of true beliefs, *we* can nonetheless know that the NED-worlders are using processes that are intrinsically reliable. We are also able to identify with fairly high precision which of the processes among our cognitive faculties that are disposed to produce a high ratio of true beliefs based on how we experience our own use of them. It seems that regardless of how we approach the metaphysics of dispositions, we should be able to confer the dispositional property of being reliable to the NED-worlders' processes insofar as they are using the same processes that we are using to acquire beliefs about the world. Seeing as the internalist posits that there must be something that confers justification on our beliefs as well as the NED-worlder's beliefs and that whatever does so must be identical in both circumstances, the explanation for this could be that their belief-acquiring processes possess the same type of intrinsic dispositional property.

Similarly, we are often able to identify the maskers of the intrinsic dispositional property of being reliable when, despite using processes with this property, we are left confounded. Maskers of our perceptual processes that are also unidentifiable on the other hand are very rare and need not necessarily factor into the analysis of knowledge (but of course this can be done). Perhaps we simply *should not* include such factors into the analysis of justification or knowledge, on pain of a far too demanding and unrealistic infallibilistic notion of knowledge seeing as we obviously cannot gain knowledge about per definition unknowable deceivers.

So, in order to be epistemically justified, it may very well be enough to use whatever cognitive processes are available to you. Importantly, this is not incompatible with the idea that this is in virtue of the intrinsic dispositional reliability these processes possess. The contention here is that this is precisely the reason for why an appeal to one's internal states can be associated with epistemic justification to begin with. Were these internal states not disposed to yield a high ratio of true beliefs, they would not afford a subject the status of being justified simply in virtue of being accessible to a cognizer.

If we only consider the internal aspects of a person (their perceptual organs), we run the risk of being in unsuitable or deceptive environments that could threaten the overall truth-conduciveness of the external and internal state we are in. But is this a problem? Can internal reliabilism be enough to account for both justification and knowledge (when it is coupled with a true belief)? Ultimately this seems to come to down whether one can acquire a true belief by using processes that are intrinsically disposed to generate a high ratio of true beliefs without being in a state of knowing. Normally, environments that are highly deceptive or unfavorable only engender false beliefs, and so such examples cannot serve as counterexamples to any standard JTB analysis of knowledge involving intrinsic dispositional properties. But what about barn examples, where one acquires true beliefs through sheer luck by forming the belief 'this is a barn' about the only real barn in a field of barn-façades? The types of scenarios where the process used is clearly generally reliable yet produces an accidentally true belief due to a deceptive environment are quite difficult to handle with this conception of justification. While the accidental reliability of clairvoyance is problematic insofar as it shows that a process is reliable for reasons that have nothing to do with the intrinsic properties of the process itself, the accidental nature of the knowledge one gains of the fact that there is a real barn among many barn facsimiles is due to broader safety considerations. Consider Williamson's (2000, 128) formulation of a safety condition for knowledge (for heuristic purposes understood topologically where α and β are situations similar enough to each other, as it is in the barn scenario where the percepts are similar enough): For all cases α and β, if β is close to α and in α one knows that C obtains, then in β one does not falsely believe that C obtains. Knowledge analyzed into a JTB theory where justification takes the form of intrinsic dispositional properties of being reliable are seemingly unable to handle the barn scenario. If reliability were a mix of extrinsic and intrinsic factors, the barn scenario would simply be ruled out as a case of knowing since visual perception in fake barn county is not exactly reliable. But visual perception is intrinsically disposed to be reliable, and so we have a case of knowledge without meeting the safety condition.

A potential way forward here would be to consider Goldman's (1976) solution where a necessary condition for knowledge, over and above using a perceptual mechanism (or any intrinsically reliable process), would be to add a condition stating that there cannot be any (ibid., 786) "relevant counterfactual situations in which the same belief would be produced via an equivalent percept and in which the belief would be false." Such a condition would rule out fake barn counties and stopped clock cases of knowing. But seeing as reliability understood as a mix of extrinsic and intrinsic factors eliminates such cases by type-individuating the environment along with the perceptual process, in the process ruling out perception-in-barn-county or telling-the-time-on-a-stopped-clock-scenarios as cases of knowledge in that they lack reliability, it may be hard to see how an intrinsic dispositional property view of reliabilism with a 'no relevant counterfactual situations' clause would be preferable to standard reliabilism. If they are judging equivalently in relevant problematic scenarios, as they seem to be doing, the dispositional view of reliabilism is perhaps still preferable as it avoids the NED-problem and it has the ability to solve certain cases of accidental reliability. As always, as some problems are solved, others arise, and so perhaps it is best to leave a more thorough discussion of the potential problems of intrinsic dispositional reliabilism for another paper. It is nonetheless important to note that if we focus too much on the subject-internal in the notion of justification, problematic results may arise that need to be addressed.

I will now proceed to conclude this paper with some clarifying remarks regarding the notion of an 'internally reliable' process and how such a position relates to externalism and internalism of justification more generally, as well as make some comments on the value of justification conceived as an intrinsic dispositional property, and, finally, its relation to naturalist conceptions of knowledge.

5. Concluding Remarks

While this account of justification also faces some problems, it is an interesting result for the following reason. If nothing else, it shows that we can accept the internalist intuition wholesale while remaining externalists about justification. The NED-problem is not a knockdown argument against reliabilist conceptions of knowledge and justification. But what can we say about the suggestion that reliability is a form of truth-conducive capability ascribable to the intrinsic properties of belief-acquiring processes?

To say that reliability is an intrinsic dispositional property does not make it an internalist notion. Whether an intrinsic dispositional property of some cognitive faculty is truth-conducive is ultimately only evaluable based on factors external to the mental content of that cognizer. In extreme cases of cognitive decline, for instance, one is not in a position to evaluate whether one's faculties are still truth-conducive. This does not mean that in most cases, in normal or ideal conditions, one is not in a position to evaluate the state of one's cognitive faculties with adequate degree of accuracy (for example, people notice if they get something in their eye, precluding them from seeing clearly, and so on). In any case, it is hopefully clear that justification in this view is not internalist, but merely a *subject-internal property*.[8]

Yet this view has some conciliatory value. The central internalist epistemic desideratum (accessibility) can likely, if not be completely accounted for, at least be appealed to, and be afforded plausibility despite insisting on truth-conduciveness as essential for epistemic justification. For instance, one can still maintain that whether we are justified is largely accessible to us (even implicationally so, with the right kind of caveats). Only in this case, it is in virtue of reliability being a property of cognitive faculties which we happen to be consciously aware of in the process of using them (in a broad sense, we know, or are in a position to know, that we are using our visual system when reading, our olfactory system when noticing a scent, etc., and we also know that these are normally reliable). If we maintain that justification is a subject-internal property, this explains the notion of accessibility as an epistemic desideratum in that we normally have some access to our subject-internal states.

Externalism need not imply that justification depends on factors external to the cognizer in a way that by necessity factors the environment into our analysis of justification. Instead, we could say that justification depends on whether one's faculties have certain epistemically valuable properties, such as being disposed to produce a high ratio of true beliefs. Madison (2021) questions the value of reliability as an intrinsic dispositional property. Justification has to have an instrumental value, in the sense that it leads us to truth. The intrinsic dispositional property of being reliable, however, is not a perfect path to truth – the path may be obstructed through various means. So, why would it be epistemically valuable? Here is my answer: Reliable processes do not necessarily yield true beliefs (the reliabilist does not demand a perfect truth ratio) but they nonetheless do so contingently[9] as long as we are in good cognitive health and use the processes in

[8] Similar views are expressed by Mulnix (2013), who notes that invoking 'mental states' or the internal properties of a subject is not a rejection of externalism. Similarly, being an internalist does not necessarily mean that the accessibility desideratum pertains to subject-internal properties (e.g., one could be said to be accessing universals or sense data; see Mulnix 2013, 37, also Fumerton 1995, 60-66).

[9] I separate "contingently" from "accidentally" here, even though they are often used

the kinds of environments we normally inhabit. We only need to include the latter in the analysis of justification if we are aiming for the guarantee that if one is justified, one is tracking truth infallibly. But, as we likely have to accept, there are no such external guarantees – at least not in the actual world.[10] As epistemically valuable states or virtues go, then, it seems that reliability as an intrinsic property of our cognitive faculties may be about as good as it gets. In any case, the fact that there are possible worlds wherein certain dispositional properties are perpetually blocked from manifesting does not mean that these are less valuable for inhabitants of worlds wherein they are not systematically precluded from manifesting themselves. A process which allows us to produce true beliefs in virtue of its intrinsic properties is valuable for precisely that reason.

Something can now be said about responsibility and its relation to epistemic justification, as well. We can follow Mulnix (2013, 47) in denying that responsibility fully exhausts the concept of justification. The similarities between the NED-worlders and us do not have to be that the acquisition of beliefs proceeds responsibly in both worlds. It may lie in the properties of the types of processes being used, therefore we do not have to invoke the notion of responsibility at all or accuse the internalist of conflating responsibility, blamelessness, and epistemic justification.

A desideratum of externalism is that a theory justification should be amenable to naturalization. Is intrinsic dispositional reliability compatible with the naturalization of knowledge? It is not entirely obvious. Insofar as justification here is a dispositional property, we should ask, can this dispositional property be a natural kind? It seems plausible enough. Kornblith's (2002, 61f.) take is that natural kinds are stable homeostatic clusters of properties and as such can factor into causal explanations or inferences based on natural laws. He claims that knowledge is precisely this type of well-behaved category (ibid., 62-63):

> The knowledge that members of a species embody is the locus of a homeostatic cluster of properties: true beliefs that are reliably produced, that are instrumental in the production of behavior successful in meeting biological needs and thereby

interchangeably. Whether one is privy to the truth is not only a matter of one's internal states, but it also depends contingently on whether the environment is perceivable and whether one is not precluded from exercising one's perceptual capacities by external influences (intoxication, brain damage, blindfolds, systematic deception by evil demons or barn-builders, etc.). The relation between justification and truth is not accidental, the property is let's say 'designed' to, or has a 'proper function,' to produce a high ratio of true beliefs.

[10] As some argue, "fitness beats truth" (Prakash, et al, 2021). Meaning we are not evolved to know, but to survive, and so we cannot take our cognitive faculties as genuinely truth-conducive.

> implicated in the Darwinian explanation of the selective retention of traits. The various information-processing capacities and information-gathering abilities that animals possess are attuned to the animals' environment by natural selection, and it is thus that the category of beliefs that manifest such attunement-cases of knowledge-are rightly seen as a natural category, a natural kind.

Kornblith, however, says knowledge is an ecological kind, consisting of a certain fit between organism and environment. While I agree with this, I think IDR can specify the way in which knowledge can be taken as an ecological kind by separating the internal justificatory aspect from the external (truth) aspect of knowledge. If knowledge is a fit between environment and organism, I take justification to be the internal aspect of this fit. Specifically, justification is equivalent to the properties of organisms that allow them to receive information about their environment. The external part is simply "truth," or the state of the environment at the time of using a reliable process. Given the full analysis of knowledge as a true justified belief, we seem to have the two parts that make up the fit between organism and world in the way Kornblith aims for. On the one end, the organism is using cognitive processes that are disposed to produce a high ratio of true beliefs, on the other, the environment lays before the organism using this process. It seems to me that an organism using such a process – without overt deception going on – would indeed acquire knowledge about its environment. Justification, one could say, is an openness to the world (as Merleau-Ponty often notes regarding perception, cf. 2012, 17). Knowledge arises when the world is not such that it would block or manipulate this openness to the world.

Seeing as there is at least one version of reliabilism that solves the NED-problem as well as the clairvoyance problem, it can at least be concluded that hope is not lost for the externalist. With this view of justification, it becomes possible to avoid relativizing reliabilism to specific types of worlds or conditions. It amounts to a notion of justification simpliciter; a notion of justification applicable in all possible worlds, one that arguably maintains an externalist spirit while heeding the internalist intuition.

This approach is also in line with Graham's (2014) arguments against transglobal reliabilism. He argues that reliabilism need not hold in all, or even most, possible environments in order to amount to justification, or (ibid., 533): "Organisms with more stable predictable natural environments can get by without such learning mechanisms; organisms do not always need transglobally reliable processes to successfully navigate their normal environments." Maybe non-accidental local reliability is good enough and transglobal reliabilism may be too much to ask since regardless of the type of cognitive process we conceive of, we can always conceive of a world in which such a process is not reliable. I believe this can be taken as further support for the view that intrinsic dispositional reliabilism is what confers justification to a subject or belief. Even if cognitive processes are sometimes systematically blocked from manifesting themselves, these scenarios are not often faced in the actual world, and so should not deter us from viewing cognitive processes that give us information about our environment as ways of acquiring justified, and in most cases, true beliefs.

References

Armstrong, D. M. 1993. *A Materialist Theory of the Mind* (Rev. ed). Routledge.

Baysan, U. 2017. "A New Response to the New Evil Demon Problem." *Logos & Episteme* 8(1): 41-45.

Bird, A. 1998. "Dispositions and Antidotes." *The Philosophical Quarterly* 48: 227–234.

Cohen, S. 1984. "Justification and Truth." *Philosophical Studies* 46(3): 279-95.

Comesaña, J. 2002. "The Diagonal and the Demon." *Philosophical Studies* 110(3): 249-266.

Fumerton, R. A. 1995. *Metaepistemology and Skepticism*. Rowman & Littlefield Publishers.

Goldman, A. 1976. "Discrimination and Perceptual Knowledge." *Causal Theories of Mind* 174.

———.1988. "Strong and Weak Justification." *Philosophical perspectives* 2 : 51-69.

———.2021. "Reliabilist Epistemology." *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/reliabilism/>.

Graham, P. 2014. "Against Transglobal Reliabilism." *Philosophical Studies* 169(3): 525-535.

———.2016. "Against Actual-World Reliabilism." In *Performance Epistemology: Foundations and Applications*. Oxford University Press.

Johnston, M. 1992. "How to Speak of the Colors." *Philosophical studies* 68(3): 221-263.

Kornblith, H. 2002. *Knowledge and Its Place in Nature*. Oxford University Press.

Lehrer, K, & S. Cohen. 1983. "Justification, Truth, and Coherence". *Synthese* 55(2): 191-207.

Madison, B. J. C. 2021. "Reliabilists Should Still Fear the Demon." *Logos & Episteme* 12(2): 193-202.

Majors, B. & S. Sawyer. 2005. "The Epistemological Argument for Content Externalism." *Philosophical Perspectives* 19: 257-280.

Mellor, D. H. 1974. “In Defense of Dispositions.” *The Philosophical Review* 83(2): 157–181.

Merleau-Ponty, M. 2012. *Phenomenology of Perception*. Routledge.

Mulnix, J. W. 2013. “Reliabilism and Demon World Victims.” *Tópicos (México)*, (44): 35-82.

Prakash, C, K. Stephens, D. Hoffman, M. Singh, M, and C. Fields. 2021. “Fitness Beats Truth in the Evolution of Perception.” *Acta Biotheoretica* 69(3): 319-341.

Williamson, T. 2002. *Knowledge and Its Limits*. Oxford University Press.