



# A Challenge for Indexical Reliabilism

Balder Edmund Ask Zaar<sup>1</sup>

Received: 30 November 2022 / Accepted: 10 April 2023 / Published online: 25 April 2023  
© The Author(s) 2023

## Abstract

The new evil demon problem amounts to a difficult challenge for the externalist about epistemic justification. Many solutions to the problem have been proffered in the almost 40 years since its first appearance in the literature. Among the more promising responses is indexical reliabilism, a combination of two versions of actual world reliabilism where “actual” denotes either the world of utterance or a rigidly determined actual world. This paper does three things. First, it attempts to clarify indexical reliabilism and how it purports to solve the new evil demon problem. Second, it attempts to mitigate some of the prominent criticism that has been leveled against the theory. Third, it poses an explanatory challenge for the theory which remains even after all of the premises supporting indexical reliabilism are accepted. The conclusion is that indexical reliabilism is not tenable until a linguistic mechanism for the use of “actually reliable” has been offered that explains how the theory avoids collapsing into a two-concepts response to the new evil demon problem.

**Keywords** Indexical reliabilism · New evil demon problem · Actualist reliabilism · Counterexample to reliabilism

## 1 Introduction to the New Evil Demon Problem

The new evil demon (NED) problem poses a serious challenge to the reliabilist project and potentially to externalist approaches to justification and knowledge in general. The problem seems to uncover a defect inherent to the idea that justification is something that obtains depending on factors external to the accessible mind contents of cognizers. Essentially, the problem purports to show that reliability is not a necessary condition for epistemic justification.

---

✉ Balder Edmund Ask Zaar  
balderaskzaar@gmail.com

<sup>1</sup> Department of Philosophy, Lund University, Lund, Skåne, Sweden

The NED problem first appeared in Lehrer and Cohen (1983) and Cohen (1984),<sup>1</sup> and it asks us to imagine a possible world that is just like ours, but where the perceptual systems that are reliable in the actual world have been made wholly unreliable by a malevolent demon. Nonetheless, all epistemic behaviors are kept the same in the inhabitants of the two worlds. Furthermore, the inhabitants of the NED world would, by stipulation of the NED example, have no way of acquiring knowledge of the evil demon's manipulation and so would view their epistemic behavior as justified (just as we do). Lehrer and Cohen (1983, 192-93) summarize the scenario, as well as the potential defect it uncovers, effectively:

Imagine that, unknown to us, our cognitive processes, those involved in perception, memory and inference, are rendered unreliable by the actions of a powerful demon or malevolent scientist. It would follow on reliabilist views that under such conditions the beliefs generated by those processes would not be justified. This result is unacceptable. The truth of the demon hypothesis also entails that our experiences and our reasonings are just what they would be if our cognitive processes were reliable, and, therefore, that we would be just as well justified in believing what we do if the demon hypothesis were true as if it were false. Contrary to reliabilism, we aver that under the conditions of the demon hypothesis our beliefs would be justified in an epistemic sense. Justification is a normative concept. It is an evaluation of how well one has pursued one's epistemic goals. Consequently, if we have reason to believe that perception, for example, is a reliable process, then the mere fact that it turns out not to be reliable, because of some improbable contingency, does not obliterate our justification for perceptual belief. This is especially clear when we have good reason to believe that the contingency, which, in fact, makes our cognitive processes unreliable, does not obtain.

So what is the problem that the new evil demon scenario presents to the reliabilist? First, reliabilism fails to explain what will be called the "internalist intuition," which is the intuition that if two subjects use the same methods to acquire their beliefs, then they will also be equally justified in holding those beliefs. Second, and perhaps more importantly, it entails a fatal contradiction for standard reliabilism (by way of the NED problem).

While the validity or reliability of philosophical intuitions may be questioned, it seems maximally charitable towards the internalist to, for the sake of argument, accept the internalist intuition and see whether it in fact entails that one cannot be an externalist about justification. In any case, the NED world inhabitants are then said to hold justified beliefs since they are just as epistemically responsible as the cognizers of the actual world (and we, of course, want to view our own epistemic behavior as justified). As Cohen emphasizes (1984, 282): "My argument hinges on viewing justification as a normative notion. Intuitively, if S's belief is appropriate to the available evidence, he is not to be held responsible for circumstances beyond his ken." If intuition tells us that the NED world inhabitants cannot be held responsible for the whims

---

<sup>1</sup> See also Pollock (1984) for a similar brain-in-a-vat example and Foley (1985) for another anti-reliabilist demon-world example.

of the evil demon and so remain justified in holding various perceptual beliefs about their world, this cannot be accounted for within the externalist theories of justification since there it is up to the whims of nature whether a given belief-forming process is reliable.

The example also engenders a strong argument against reliabilism which will be used throughout this paper to test the plausibility of a given response to the problem. The full *reductio* argument runs as follows:

1. The NED world inhabitant 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' beliefs are as justified as our own (internalist intuition).
4. Therefore, the perceptual beliefs<sup>2</sup> of the NED world inhabitants have been produced by reliable processes (1–3; 1 and 4  $\Rightarrow \perp$ ).

So not only does reliabilism need to explain why it is that we (those of us that do harbor this intuition) consider NED worlders as justified as us despite the fact that their processes are unreliable, but it entails a contradiction since perceptual beliefs are invariably unreliable in the NED world. In order for reliabilism to remain conciliatory and possibly convince the internalist to change sides, it seems that the reliabilist needs to (1) provide an explanation for the intuition that NED world inhabitants are as justified as we are. In order for reliabilism to be tenable, the reliabilist (2) needs find a way to mitigate the *reductio* argument.<sup>3</sup> Given that we are forced to accept the internalist intuition (which, ultimately, we may not be), internalism is *prima facie* favorable in that it is seemingly unfazed by the NED problem (although some argue that internalists face the problem as well, see, e.g., Moon, 2012). Internalism – in the broadest of strokes — states that justification is something found in the accessible parts of the minds of cognizers, so the inhabitants of the actual world are precisely as justified as the inhabitants of the NED world since they share all non-factive mental states. Kornblith illustrates this anti-reliabilist and internalist point of view very well in the following quotation (1983, 45):

Since epistemically responsible action may result in something less than reliably produced belief, an agent may be justified in holding a belief without that belief being reliably produced. Beliefs produced by unreliable processes, where the extent of the unreliability would not be detected by an epistemically responsible agent, are nonetheless justified.

We now seem either to be forced to accept internalism or to advocate for a theory of justification that involves two concepts that accommodate the internalist intuition,

<sup>2</sup> Or any other kind of belief we paradigmatically take to be justified or reliably acquired (so instead of a perceptual belief, it could be one arrived at through using sound reasoning, and so on).

<sup>3</sup> For one can do the former without doing the latter, and vice versa.

on the one hand, and the externalist intuition (roughly, that justification has something to do with truth in a way that is not always accessible to us), on the other. A two-concept response is for instance proposed by Lehrer and Cohen (1983), and some prominent proponents are Goldman (1988) and Sosa (1993 & 2001 & 2003, 157). According to Lehrer and Cohen, the NED problem leads to the necessary bifurcation of justification into so-called verifical and personal justification. The former is what one has if one's faculties are unmanipulated by extrapersonal forces (i.e., no processes that have been made unreliable by way of a demon, an illness, the whims of nature, or whatever else we can conceive of). Personal justification, according to Lehrer and Cohen, depends solely on the coherence of one's personal belief acceptance system.<sup>4</sup> Prima facie, we can say that the internalist can account very well for personal justification, in a way that the externalist cannot. We should, with the same amount of emphasis, be inclined to ask whether the internalist can account for verifical justification. Arguably, it is the unification of these perspectives on justification that the NED world problem invites us to consider.<sup>5</sup> Alston (1993, 534) can be viewed as urging us to consider something along the same lines. It is not only (and for him it is not at all) a matter of finding a *correct* notion of justified belief, but a matter of weighing epistemic desiderata:

Instead of trying to decide what is the correct account of justified belief, the real problem is one of understanding and interrelating the various epistemic desiderata, determining which of them are feasible goals and what it takes to achieve each of those feasibilia, and identifying the contexts (interests, aims, problems) for which one or another is most important.

This approach seems virtuous and valuable in its own right, perhaps even necessary in order for the externalist and internalist to avoid simply denying the importance of the desiderata of their "opponent." However, if either externalism or internalism can account for each relevant desiderata with a single concept of justification, this would surely be preferable. To deviate from Alston, "correctness" should in my view not be read as just being about positing theories of justification that avoid counterexamples, but about appealing to and meeting relevant desiderata in the precise manner in which Alston describes. He is likely correct in thinking that most philosophers would not deny the importance of the most commonly discussed epistemic desiderata (i.e., epistemic responsibility, cognitive accessibility, coherence, truth-conducivity, groundedness or "basing relation"), but it seems that

<sup>4</sup> They make this distinction very clear in the following quote: "Personal justification, by contrast, depends only on coherence with the acceptance system that one actually has. Hence, a person in the demon example is personally justified in his perceptual belief but is not verificaly justified. In this case, one is personally justified because the belief is probable relative to the assumptions one makes about one's reliability under the conditions of perception, but one is not verificaly justified because it is not probable relative to the correct assumptions about one's reliability under the conditions of perception."

<sup>5</sup> Alston (1993) frames the NED problem as a conflation of two desiderata, the basing relation and truth-conducivity, which throw us between deeming the NED worlders justified in view of the basing relation desideratum since the beliefs are based on the same kinds of processes, but unjustified in view of the truth-conducivity desideratum since the processes are no longer truth-conducive.

we should be inclined to subscribe to the theory of justification which can best meet these desiderata. Although, importantly, not necessarily by making each desideratum an element in a set of necessary and sufficient conditions for the concept of justification.

Many difficult questions face the reliabilist. Can the reliabilist effectively respond to the NED problem in a way that explains away or obviates the concerns of the internalist intuitions without thereby being forced into internalism? We also have to ask, with Williamson (2016, 3), whether the NED world inhabitant is justified, or merely blameless (also discussed in Pryor, 2001 and Ball & Blome-Tillman, 2012). Of course, we cannot fault the NED inhabitant for what is outside of their frame of possible comprehension and control, but it is not obvious that faultlessness or blamelessness is the same thing as justification. Some theories of justification (for instance in Alston, 1985) amount to blamelessness (as well as Goldman's "weak justification"). However, it is not clear that blamelessness has anything to do with being epistemically justified. Being justified may involve being blameless but being blameless may not be sufficient for justification. That is, blamelessness does not by necessity have to do with using particular types of behaviors or anything that has to do with justified epistemic behavior. Blamelessness, minimally, involves not being to blame, but this does not mean one has done anything in order to be viewed as being justified since it could simply involve absence of irresponsible behavior. That is, epistemic blamelessness seems to be negative in nature (that one has *not* done something wrong), whereas justification involves having done something right. Thus, it is possible to frame justification as a matter of using a set of particular types of behaviors in virtue of them having certain properties (such as truth-conduciveness). In any case, there is no obvious way in which to equate blamelessness with justification.

The reliabilist thus stands before a rather difficult task. Cohen (1984) aptly emphasizes the need for epistemologists to clarify the truth-connection involved in genuine states of knowing and similarly poses the question as to whether justification is necessarily the same as being objectively (in the sense of "determined by external factors") connected with truth. The question then, more specifically, is the degree to which justification must be a truth-conducive phenomena. The reliabilist must also consider what can be done to avoid having epistemic justification be conflated with other concepts such as responsibility or blamelessness, which do not, at least prima facie, have anything to do with truth. At the same time, reliabilism must explain how an externalist theory of justification can account for these more strictly normative notions that are undeniably related to being justified, although perhaps peripherally.

What kind of relation can there be between justification and truth? Lehrer and Cohen (1983, 192) present a few ways in which truth and justification can be related that all end up being untenable as far as they are concerned.

The relation, they claim, can be logical, where if a belief is justified, it is also true. Such a view inevitably leads to skepticism, however, because there is no way

to establish that if one uses all perceptual or intellectual powers available, one is guaranteed to obtain a true belief.<sup>6</sup> Some information-based theories of justification, as well as causal theories of knowledge, like those of Dretske (1981) and Armstrong (2000), respectively, on the other hand, can be seen maintaining that the relation between truth and justification is nomological. In the latter case, we simply see that no such law exists since we oftentimes have justified beliefs, and so think we have knowledge, but nonetheless, the belief turns out to be false. The relation between truth and justification, it is said, must maintain the possibility of error. In other words, *the failure of a justified belief to be true cannot be a violation of either a logical or nomological law*. This is precisely what reliabilism amounts to. It does not have the strictness of a logical or nomological relation, but still informatively explicates the relation between truth and justification; a relation made up of processes that reliably, but not invariably, produce true beliefs. One could perhaps even frame reliabilism as a kind of fallibilist-nomological relation between truth and justification — given the absence of certain conditions that invalidate our perceptions or methods, one gains knowledge by using reliable (truth-conducive) belief-acquiring processes. This fact, coupled with other values of reliabilism, such as its aptness for naturalization and its compatibility with animal knowledge (which makes for a broader conception of knowledge), serves as more than enough motivation to attempt to find a solution to the NED world problem as well as find a version of reliabilism that can meet the constraints of the internalist intuition.

Ideally, then, the solution to the NED problem would be one that is maximally conciliatory; that is, it would heed the internalist intuitions without thereby conceding too much and thereby losing the many virtues inherent to externalist conceptions of knowledge and justification.

Comesaña's (2002) indexical reliabilism meets these constraints directly. It is a version of actualist reliabilism that takes advantage of Stalnaker's two-dimensional semantics in order to basically maintain a single concept (taken in a very particular sense, as will be explained below) of justification while exploiting the pragmatic mechanism of "actually" to cleverly avoid the two main types of counterexamples to reliabilism. That is, the NED problem as well as the issue actualist reliabilism has with accounting for non-actual world reliability where processes that may not be reliable in our world can be reliable in other worlds without thereby conferring justification on beliefs acquired via those methods simply in virtue of it occurring in the wrong world. Indexical reliabilism also explains how the NED worlder can acquire beliefs unreliably in one sense, yet still hold justified beliefs in virtue of the belief being "actually" reliably acquired. It faces some striking criticism, however, so the

---

<sup>6</sup> Of course, on knowledge-first accounts (e.g., Williamson, 2002), or where justification is identified with knowledge, truth is entailed by justification as long as knowledge is factive. However, this does not mean that if one engages in paradigmatic knowledge-acquiring process, that one will acquire knowledge. On Williamson's view, seeing that P entails that P, but attempting to see that P obviously does not logically entail that one sees that P. The latter is more what I have in mind here. Even though it is possible to make truth and justification conceptually related, there remains a problem as to whether any given case of seeming perception is genuine (factive) perception.

aim of this paper will be to evaluate the plausibility of indexical reliabilism given the criticism put forth primarily by Ball and Blome-Tillmann (2012). As the title suggests, even with a defense of indexical reliabilism against this recent criticism, it struggles to provide a satisfying solution to the NED problem, and so the indexical reliabilist is facing a rather significant explanatory challenge.

Let us now proceed with an account of indexical reliabilism and see how it solves the NED problem.

## 2 Indexical Reliabilism

Comesaña's indexical reliabilism is a theory that amounts to an argument that the NED worlders are in fact justified in a way that can be explained by modifying standard reliabilism — somewhat significantly — while maintaining that there is essentially only one kind of concept involved whenever a belief is justified. If there is a version of reliabilism which makes the result of the NED argument (1–4 in the introduction) acceptable as well as explain the internalist intuition, this would presumably resolve the issue. Naturally, one has to ask what could make it true that the NED world inhabitants have formed their beliefs by way of reliable processes when it is simultaneously stipulated that their processes are unreliable. As the actual-world responses to the NED problem suggest, it is the addition of “actual” that make a formulation of reliabilism acceptable even in the face of the NED argument — but only one of the possible interpretations of “actual” can help solve the NED problem. As Lewis (1970, 185) has noted, “actual” has more than one sense. The possible world that “actual” denotes depends on which world is under consideration in the context of utterance. Specifically, the use of “actual” can either pick out the world in which a sentence containing the word was uttered (the non-rigid reading) or it refers to one and the same world picked out by any uttered sentence containing the word “actual” in any possible world (the rigid reading).<sup>7</sup> In the former case, the world is designated through the utterance; in the latter, it has been (rigidly) designated to be one and the same world regardless of context of utterance.

Comesaña (2002, 258), taking the two readings of “actual” as his starting point, comes to the following two formulations of actualist<sup>8</sup> reliabilism:

**Rigid Reliabilism:** A belief is epistemically justified if and only if it was produced by a process that is actually reliable.

<sup>7</sup> The different readings of “actual” are called “rigidified” and “unrigidified” in *Plurality of Worlds* (Lewis, 1986, 94). The following quote illustrates the difference further (ibid.): “So you say ‘Yesterday it was colder than it is now’, and even in the scope of the time-shifting adverb, ‘now’ still refers to the time of utterance. Likewise you say ‘Yesterday it was colder than it is at present’, and the reference of ‘present’ is unshifted. But if you say ‘Every past even was once present’, then the time-shifting tensed verb shifts the reference of ‘present’. I suggest that ‘actual’ and its cognates are like ‘present’: sometimes rigidified, sometimes not. What if I’d had an elder sister? Then there would have been someone who doesn’t actually exist. (Rigidified.) Then she would have been actual, though in fact she is not. (Unrigidified.) Then someone would have been actual who actually isn’t actual. (Both together.)”

<sup>8</sup> “Actualist” here is with the rigid and non-rigid distinction in mind, and so “actualist” denotes both varieties of actualist reliabilism.

Non-rigid Reliabilism: A belief is epistemically justified if and only if it was produced by a process that is reliable in the world where it is used.

Together these capture the idea that when we say that someone is justified in the actualist reliabilist sense (or that one of their beliefs are actually justified), this can, depending on the context, either mean that they hold a belief that was acquired through a process that is reliable in the actual world (that is, our world) or it can mean that it is a process that is reliable in the world where “S is justified in believing that P” has been uttered. This dual interpretation is denoted by “actually” in the first line of the argument below. Let us now review the argument from the introduction with the modified version of reliabilism:

1. The NED world inhabitant cannot acquire beliefs reliably (NED world stipulation).
2. A belief is justified if and only if it has been formed by way of an actually reliable process (indexical reliabilist assumption).
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 produced by actually reliable processes (1–3).

How does the potential dual interpretation of reliabilism change the outcome of the argument? Given the interpretation of “actually” along the lines of rigid reliabilism, the conclusion is no longer a contradiction, since the beliefs of the NED world inhabitants have in fact been produced by processes that are reliable in our world. They just happen to not be reliable in the world of the evil demon. This also explains the intuition (3) somewhat, since it is quite natural to think that what is reliable to us will be reliable to those just like us, that is, those with the same perceptual systems as us (even if those happen to be operating in extraordinary circumstances). While this version of reliabilism solves the NED problem, it brings on new counterexamples it is unable to solve.<sup>9,10</sup> Moreover, as Comesaña writes (*ibid.*, 261), it is non-rigid reliabilism that is necessary for knowledge. That is, for any given world, if we are looking to acquire knowledge in that world, we are concerned with the processes that yield a high ratio of true beliefs *in that world*. Non-rigid reliabilism, on the other hand, is the interpretation of actualist reliabilism that engenders the NED problem. As we can see if we input non-rigid actualist reliabilism as the assumed

<sup>9</sup> According to Comesaña, however, a contextual change permits a different interpretation of “actually,” which solves many of these issues, for instance, with the rigid reading we avoid the counterexample to actualist reliabilism where we are asked to imagine a possible world wherein clairvoyance is a reliable process — again, this seems to explain our intuitions as well as give us a way to plausibly claim that the clairvoyance is not “actually” reliable.

<sup>10</sup> See 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.”



theory of justification, the NED argument spits out a contradictory conclusion. From the reformulation of reliabilism into non-rigid reliabilism, we can derive that the perceptual beliefs of the NED world inhabitants have been produced by processes that are reliable in their world (this is, again, necessitated by the intuition that they are as justified as we are in holding various perceptual beliefs):

1. The NED world inhabitant cannot acquire beliefs reliably (NED world stipulation).
2. A belief is justified if and only if it has been formed by way of an actually reliable (non-rigid) process.
3. The NED world inhabitants' beliefs are as justified as our own.
4. Therefore, the perceptual beliefs of the NED world inhabitants have been produced by processes that are reliable in their world ( $\perp$ ).

Seeing as the thought experiment has as a premise that all perceptual processes in the NED world are unreliable, this clearly amounts to a contradiction. Rigid reliabilism, on the other hand, shows how there may nonetheless be a sense in which a subject may be justified without having genuine knowledge. They are still using processes that are reliable in the actual world. Moreover, it is the non-rigid version of reliabilism that solves the clairvoyance problem, since in a non-actual world wherein clairvoyance is reliable for genuine reasons (let us say there are some new natural phenomena that allow for such a process to reliably produce true beliefs), we can explain how they are justified even though they are not using processes that are reliable in our (the rigidly designated) world. We still require the dual interpretation of “actually,” then, and this is what indexical reliabilism purportedly allows for. The NED argument can then more generally be said to fail by way of the fallacy of equivocation; it does not heed the two interpretations “actually” actually permits by taking non-rigid actualist reliabilism as being the only way to interpret reliabilism.

Comesaña still has to show how we can arrive at the first interpretation in one context but at the second interpretation in another. How can one word express two different senses without this also amounting to two different concepts of reliabilism and thus a kind of two-concepts response? This is where two-dimensional semantics becomes necessary. It is a framework for representing indexical expressions like “here” and “now” and the fact that sentences with these words can express different propositions depending on the context of utterance as well as the context of evaluation. A similar framework, if applicable, can infuse “actually” with the dual readings necessary to undermine the more prominent attempts to provide counterexamples to reliabilism (i.e., it solves the NED problem and the clairvoyance problem simultaneously) while maintaining that justification essentially amounts to one kind of concept.

First, Comesaña illustrates the difference between rigid and non-rigid reliabilism with the help of two-dimensional matrices (Stalnaker calls these “propositional concepts”). Say  $W_0$  is the actual world,  $W_1$  is the demon-world. Now the vertical axis is the world in which “S is justified in holding some perceptual belief B” has been uttered, while the horizontal axis denotes the circumstance of evaluation for each

utterance. That is, the former represents how the world determines what is being said, the latter the truth value of what is said. This is far from clear, but Stalnaker (Stalnaker, 1999, 295) provides some clarification. Consider the following quote:

The two-dimensional framework was deployed, in the first instance, as a piece of descriptive apparatus for representing the way that semantic values depend on the facts. We need two dimensions since we start with the fact that the truth value of a proposition (at least a contingent proposition) depends on the facts. But since the identity of the proposition expressed in a given utterance also depends on the facts, the truth value of the utterance will depend on the facts in two different ways: first, the facts determine what is said; second, the facts determine whether what is said is true.

The vertical axis then can be said to be the way in which the facts determine what is said, whereas the horizontal axis is what ultimately determines the truth value of what is said.

Implementing Stalnaker's framework<sup>11</sup> now yields the following two-dimensional matrix for rigid reliabilism:

	$W_0$	$W_1$
$W_0$	T	T
$W_1$	F	F

*Matrix 1*

For **non-rigid** reliabilism:

	$W_0$	$W_1$
$W_0$	T	F
$W_1$	T	F

*Matrix 2*

Note now that rigid reliabilism has a diagonal (upper left to bottom right) which carries the same truth values as the horizontal truth values of non-rigid reliabilism. In Stalnaker's terminology, Matrix 2 would be a projection of the diagonal of Matrix 1, which is represented by a dagger operator (Stalnaker, 1978, 319: "The dagger is an operator which takes the diagonal proposition and projects it onto the horizontal"). We can now say that indexical propositions have a horizontal and a diagonal reading (as per the framework put forth by Stalnaker). Comesaña's single concept solution is now hopefully clear; both readings of "actually" can be contained in one and the same utterance (presumably because of the shared truth value of the

<sup>11</sup> For a perhaps more understandable or simpler implementation of Stalnaker's framework, see Stalnaker (1978, 317-318).

diagonal reading of the rigid reliabilist matrix and the horizontal reading of the non-rigid reliabilist matrix, although this is not explicit in Comesaña's text). "Actually reliable" may be said to contain two readings which when alternated appropriately, solve both the NED problem and the clairvoyance problem. So far so good. The final step is to motivate the reason for taking the horizontal reading as opposed to the diagonal, but this is where we hit a snag (or, really, many snags). Ball and Blome-Tillman (2012) represent the main opponent to indexical reliabilism. The following section will go through their criticism in order to initially defend indexical reliabilism in order to attempt making the best possible case for the theory — in terms of the details — before countenancing it with some broader concerns.

### 3 Some Criticisms of Indexical Reliabilism and How They Fail

Ball and Blome-Tillman (2012) list a number of issues with indexical reliabilism, many of which seem serious enough to warrant the dismissal of the theory. These issues are mostly related to the idea that Comesaña's indexical reliabilism fails to implement Stalnaker's two-dimensional semantics properly. This failure, in turn, leads to potential problems when it comes to the claim that there are in fact two readings of the propositional concept in the relevant circumstances. As we shall see, there are issues that face Comesaña's indexical reliabilism which do not seem to face the kind of statements that normally lend themselves to being modelled by Stalnaker's two-dimensional semantics framework. This implies, minimally, that there could be some difference between the way we view indexical terms normally and the way Comesaña construes them.

The first criticism presented is against the idea that the horizontal proposition is true while the diagonal proposition is not. Ball and Blome-Tillman do not seem to think there can be two distinct readings of "actually" in the actual world since the two readings have the same truth value in the actual world. In our context of evaluation, where the sentence is also uttered in our world, both readings are of course true. Ball and Blome-Tillman writes (*ibid.*, 8) that "Thus, if Comesaña wants our actual utterance of ["The NED-worlder's belief was produced by a process that is actually reliable"] to have two differing readings with differing truth-values in our world, then those readings cannot be the horizontal and the diagonal propositions respectively: both are, after all, true at [our actual world]."

A few things should be said in response to this. Notice, firstly, that there is a difference between the utterances chosen as examples for the two-dimensional propositions and that this difference somewhat misconstrues Comesaña's theory, making it seem as if this is a problematic result. Comesaña on the other hand uses an utterance where the relevant subject is one that exists in the world of the utterance (i.e., "Ernie's beliefs are justified," the insertion of the reliabilist formulation of justification was left to the reader), whereas Ball and Blome-Tillman use an utterance that refers to a subject in the NED world (they use the term "biv" for "brain in a vat," denoting the possible world closest to ours where brains in vats are a reality and so the relevant utterance is: "Biv's beliefs were produced by a process that is actually reliable"). Two different individuals are selected here, it seems. In the one case, we

have the same individual denoted in each utterance; in the other, the individual varies along with the possible world in which it was denoted. If we take Comesaña's example utterance instead, where the individual denoted depends on the world in which a sentence denoting them was uttered, we do in fact want both readings to be true in our world. That is, if the sentence expressing that someone is justified is uttered in our world, and we evaluate it in our world, we want both readings of reliabilism to be true in *our context of evaluation*. This is the main purpose of indexical reliabilism! Since the utterance of "actually" can designate both the world in which it was uttered and the actual world, taken rigidly, and it so happens that it was uttered in the actual world, both readings denote one and the same world, since the non-rigid and rigid reading designate the same world. Therefore, the truth values of the expressed proposition naturally converge as well. This is basically the idea of Stalnaker's third principle of assertions (1978, 325): "The same proposition is expressed relative to each possible world in the context set." If this were not the case, indexical reliabilism would fail outright. Therefore, it seems that we do achieve the desired effect, after all, if we insert Comesaña's intended utterance and relativize the utterance of "S is justified in believing that P" to subjects of that world. The truth values are the same, but the mechanism that picks out the worlds which make the utterances true still differ.

So, the two readings remain, it just so happens that they are logically equivalent readings in our context of evaluation (the actual world taken rigidly as our world). When we consider the entire matrix, we can still see that there is a difference, as was accounted for above. If the utterance is made in our world, "justification" denotes processes that are reliable in our world according to rigid reliabilism, and so "S is justified in believing that P" comes out true in the horizontal reading in both our world and the NED world. But since non-rigid reliabilism (and so the diagonal reading) specifies what is said in terms of the NED world, along with the facts of the NED world, it naturally comes out false that S is justified in believing that P if this is uttered in the NED world (where both the individual and their perceptual processes are non-rigidly designated). In the world where the process took place, the process is not reliable. The only epistemic justification that can exist in the NED world is one that is evaluated based on facts regarding the actual world, and this is how the indexicalization of reliabilism avoids the NED problem. In any case, it seems that Comesaña's matrices work as intended in this regard. The NED problem requires us to consider at least two worlds as having at least two different functions (once as a determiner of what is said, second as a determiner of whether what is said is true), just as the two-dimensional semantics forces us to do, and so the analysis of the propositional concept is incomplete if it is based on the readings of the actual world alone. The two readings are perfectly accessible once the full conversational context, involving two contexts of utterance as well as two contexts of evaluation, is included.

Another aspect touched on by Ball and Blome-Tillman that needs to be discussed is their claim that the relevant proposition is the diagonal one (2012, 1322). They quote Comesaña as saying that the diagonal proposition<sup>12</sup> is false in the actual world. However, Comesaña's point is in fact that that "demonic reliability" ("the beliefs of demons were produced by a process most of whose outputs would be true," Comesaña,

<sup>12</sup> Namely (ibid., 1321): "Ernie's beliefs were produced by a process that is reliable in Ernie's world".

2002, 255) is “false in the actual world, false *simpliciter*.” There is an understandable difficulty in modelling this statement with the two-dimensional semantic matrices. *Demonic reliability*, I believe, should be interpreted as an utterance in the actual world that is evaluated based on the facts of the NED world. What is said is determined by our world, then, which means that the reliable processes denoted are the ones that are reliable in our world. These, as stipulated, are not reliable in the NED world, and so the statement made in our world that the beliefs of the demoners were produced by processes that are reliable to us is indeed false. This is reflected in the top-right of the horizontal reading of the non-rigid reliabilist matrix which corresponds to the bottom-right of the diagonal reading of the rigid reliabilist matrix. It turns out this has no bearing on the argument Comesaña is making, which is merely the idea that the propositional concept of rigid reliabilism can be viewed as harboring the relevant horizontal reading of the propositional concept of non-rigid reliabilism.

But according to Ball and Blome-Tillman, there are more ways in which Comesaña’s use of two-dimensional semantics does not cohere with Stalnaker’s account of them. For instance, it is stressed that both readings need to be available at the same time given the context of the NED problem; otherwise, one cannot claim that the NED argument is merely a case of fallacy of equivocation. They hold that according to two-dimensional semantics, however, it is *usually* not the case that both the horizontal and diagonal readings of a propositional concept are available in one and the same context. Furthermore, it is quite rare that one ends up expressing the diagonal of a propositional concept — it happens in contexts where the horizontal is rejected, for instance due to it being perceived as flouting one of the Gricean maxims (Grice, 1989, 26-31). A more specific example would be if one asserts a proposition which is trivially true or necessarily true, this triggers a diagonal reading since such instances can be cases of flouting the maxim of quantity (that one should try to be as informative as possible, and thus not assert that which is presupposed to be true).

According to Stalnaker, the possible worlds which are under consideration when an assertion is made are determined by the presuppositions of those involved in the conversational context. In our case, we are explicitly aware of the possible worlds under consideration. The problem as Ball and Blome-Tillman see it is that we supposedly cannot generate the two readings in the actual world since then the horizontal reading is true in every world under consideration. This would, in Stalnaker’s framework and for Gricean reasons, force the diagonal reading and the diagonal reading alone because to assert something that is true in all possible worlds is to assert something that is already presupposed and hence to do something that has already been done through a previous assertion or that is already common (implicit) knowledge (Stalnaker, 1978, 325). In a way, one is then not asserting, but reaffirming, or something along those lines. Hence, the two readings cannot be maintained in one and the same conversational context, and thus, there is no fallacy of equivocation, and the NED problem remains unresolved.

Ball and Blome-Tillman are worried that Comesaña’s insistence on the two readings being available in the actual world breaks Stalnaker’s first principle of assertions (ibid.): “A proposition asserted is always true in some but not all of the possible worlds in the context set.” To clarify, the context set is the set of possible worlds that are compatible with what is presupposed (common knowledge of all

participants according to the speaker) in a conversational context, or alternatively, the set of “live options” recognized by the speaker to be relevant to the conversation. If the only world determining what is said by “S is justified in believing that P” was the actual world, there would indeed be a problem with asserting that which is necessarily true.

The crux of Stalnaker’s two-dimensional semantics, however, is not just the obvious idea that the truth value of a proposition depends on the facts, but that what is said is itself in part determined by the facts. If we then, again, consider two different possible worlds, we cannot simply consider the horizontal values (as functions from worlds to truth values), that is, the context of evaluation. Since the context set is determined by the utterance itself, it is necessary to consider the utterance that occurs in the NED world as well. Otherwise, we would have no reason to consider the context of evaluation of the NED world, and if that were the case, we would not be able to discuss the NED problem at all. If all we had to consider was the actual world, actualist reliability would be tautological. Simply put, we need to include the possible world’s influence on what is said (the vertical axis) in order to model an assertion involving an indexical term like “actually.” There is simply little sense in talking about a context of evaluation without a context of utterance. So when we now, as intended, take the NED world context of utterance into account, we end up with two different propositional concepts involving false utterances (meaning, there is no necessarily true proposition involved).

Another aspect potentially missed by Ball and Blome-Tillman’s critique is that Stalnaker makes the Kripkean (1981, 53-57) distinction between a priori truth and necessary truth. A statement is true a priori if it is true in every relevant context (it is thus based on a common definition of some term), whereas a necessary truth would be true regardless of any particular context set. Or as Stalnaker writes (1978, 320): “An a priori truth is a statement that, while perhaps not expressing a necessary proposition, expresses a truth in every context. This will be the case if and only if the diagonal proposition is necessary.” There is no propositional concept under consideration either in Comesaña or Ball and Blome-Tillman that meets the requirement of being true in every relevant context, however (as the two matrices show), and there is no diagonal proposition which is necessary. Stalnaker, furthermore, shows what an a priori propositional concept looks like (ibid., 320-21), and he clearly includes the context of utterance in the analysis. To illustrate, the propositional concept of the statement “This bar is one meter long,” in worlds  $i$ ,  $j$ , and  $k$  where the bar is 1 m, 2 m, and 3 m, respectively, is modeled as follows:

	$i$	$j$	$k$
$i$	T	T	T
$j$	T	T	T
$k$	T	T	T

*Matrix 3*

This matrix models the fact that what is said designates (non-rigidly) the referent of “1 m” as a function of the context of utterance and context of evaluation. There

is no equivalent a priori propositional concept under consideration in Comesaña; we can therefore not say that the diagonal proposition is forced in general, unless we are only considering the actual world. If that were so, again, we would not have to use two-dimensional semantics to model the assertion since we would not be forced to consider the possible worlds described in the counterexamples to reliabilism.

The main idea is that two readings of reliabilism, let us call them rigid and non-rigid, can be contained in one and the same propositional concept. This has not been shown to be false by Ball and Blome-Tillman. I can only conclude that Ball and Blome-Tillman's argument against Comesaña fails to consider many crucial aspects of Stalnaker's two-dimensional semantics and its relation to the modelling of the NED problem. Comesaña (2010, Footnote 24), to some extent, regards his indexical reliabilism as a novel and independent from Stalnaker's conception and implementation of his two-dimensional framework. This potentially means that even if Ball and Blome-Tillman were to be correct in their criticism, the arguments based on Comesaña's inappropriate implementation of Stalnaker's framework would be unable to decisively refute indexical reliabilism as long as Comesaña can provide alternative non-Stalnakerian explanations of how to understand the theory.

Ball and Blome-Tillman focused a lot on the details of the mechanisms of indexical reliabilism, for good reason. They are rather complicated and seem far from intuitive ways in which indexical terms work, and it is unclear how to appropriately implement the two-dimensional framework and various Gricean considerations. The following section will therefore discuss some broader concerns relating to the nature of generating the two different readings of reliabilism. Some linguistic mechanism must be involved in this process. But which linguistic mechanism could generate the two different readings? Can "actually" behave like an indexical in the way intended by Comesaña? Can we really view a term that has a rigid and non-rigid reading as unambiguous (as not denoting two different concepts)? The remainder of this paper will attempt to argue for why these questions must be answered in the negative.

#### **4 A Broader Concern with Indexical Reliabilism**

A more general concern with indexical reliabilism pertains to the question whether "actually" behaves like a genuine indexical, and whether we are not forced into a two-concepts notion of justification after all. For there seems to be a degree of arbitrariness both pertaining to the motivation of one of the readings being picked in any given scenario, as well as the motivation behind taking one of the readings to be primary. That is, why should we take rigid reliabilism to be the standard reading which need special Gricean motivations for triggering a second reading (non-rigid reliabilism), and not vice versa?

What Ball and Blome-Tillman more generally highlight is the concern regarding how one simultaneously maintains two readings of "actually" in addition to having a mechanism to trigger one of the readings when it is desirable for such a reading to be triggered. Their focus is primarily on the problems relating to the upholding of both readings in any given context of conversation. However, an equally big problem

arises when it comes to triggering one and only one reading in any given scenario which fits the picture of how indexical work. *Prima facie*, making use of indexicality seems promising since a key feature of indexicals is that their content is in part determined by the context of utterance and that this means the referents can change while the semantic underpinnings or mechanisms remain the same. However, it is not clear how the utterance of “actually reliable” leads to either of the readings of indexical reliabilism by way of the context of utterance.

If, given the discussion above, we accept that the dual readings can be maintained at all times, we nonetheless must ask what could motivate either reading. While indexical terms usually need no motivation other than the context itself in order to denote something, this does not seem to be the case for “actually” as described by Comesaña. Or, alternatively, the choice between readings could be said to be motivated by the wish to avoid the *reductio* argument which the NED problem presents to the reliabilist. This strikes me as a motivation not at all in line with how indexical terms normally function, and it does not seem to cohere with some standard accounts of indexical expressions. For instance, let us take Kaplan’s “obvious principles” of indexicals (Kaplan, 1989, 492) as an example:

Principle 1: The reference of a pure indexical depends on the context, and the referent of a demonstrative depends on the associated demonstration.

Principle 2: Indexicals, pure and demonstratives alike, are directly referential.

A short explanation of what “pure” denotes in the case of indexicals is required. Regarding pure indexicals, Kaplan writes (*ibid.*, 4919):

For [pure indexicals], no associated demonstration is required, and any demonstration supplied is either for emphasis or is irrelevant. [...] **The linguistic rules** which govern their use fully determine the referent for each context.

“Actually” is intended, presumably, to pick out a single possible world, even if its use entails the possibility of assigning two different ones. But if we split the possible interpretations of “actually” in two, without supplying the linguistic rules which govern their use, do we not lose the essential qualities that would make it an indexical? It seems that we do, since we are no longer able to say that “actually” designates a singular possible world in every instance of “actually” being used. The alternative seemingly is that we are stuck with a genuine ambiguity, involving two significantly different mechanisms of designating the possible world under consideration. For when we say, “Visual perception is an actually reliable process,” which linguistic or extralinguistic context could determine our interpretation of “actually” in the sense suggested by Comesaña? The problem is thus broader than we have seen so far, since even if we are able to maintain the two readings, in order for this approach to work, we need to also have a way to get one and only one reading in any given context in a relatively systematic or predictable manner (or perhaps it would suffice to get one of the two readings merely in the two counterexamples to the two forms of actualist reliabilism). Before discussing the linguistic mechanism involved in more detail, there is another way to approach the problem of potential ambiguity once we maintain the two readings, which is to simply accept it and claim that this ambiguity is a feature, not a bug. If there are ambivalent intuitions regarding



clairvoyance-world reliability, the fact that we have two readings and no mechanism to decide on one and only one reading would be acceptable as long as one could say that the dual-reading captures our ambivalent intuitions. This strategy seems to be at work in Comesaña's (2010, 590-592). The problem with this is that while there are cases of strange alien reliability that might sound unintuitive to us, there are also cases of alien-world reliabilism which are not as confusing. When we say that clairvoyance is reliable in some world, and leave it at that, it would perhaps seem irresponsible to engage in doxastic practices involving clairvoyance (to us). But were we to specify that clairvoyance in this world has to do with newly discovered physical mechanisms and phenomena, and through this show that the belief-acquiring process as legitimate as visual perception is in our world, indexical reliabilism would not get away with ambivalence. In such a world, it seems that we would in fact be justified using clairvoyance, and so we would have to have produce the diagonal reading and diagonal reading alone, in order for it to be true that inhabitants of that world are justified in using clairvoyance to acquire beliefs. With that hopefully made clear, I will now attempt to demonstrate the difficulty of obtaining one and one reading alone in more detail.

For it is indeed very complicated. The desired reading in each imagined scenario needs to be the one which avoids contradiction in order for indexical reliabilism to work as intended. Ball and Blome-Tillman suggests that Comesaña could maintain that the reading of "actually" in each circumstance is up to the speaker and that the reading would depend on the speaker's communicative intent. Ball and Blome-Tillman bring up the fact that nobody really maintains that the choice between propositions in this case is entirely up to us. However, I do not believe this is the problem. The problem with this strategy is not that it is an uncommon view, or that it has no backing in the literature. The problem, as I take it, is that the strategy seems to run counter not only to the idea Stalnaker is trying to get at (that what we say in part depends on the world) but to the very notion of indexicals as well. Namely, that what is said is not only determined by the intentions of the speaker or the semantics of the expressions used, but by the context of utterance (the facts) in a way that has to be modelled separately from what is semantically (conventionally) expressed by an utterance since semantics alone<sup>13</sup> cannot capture the full extent of any utterance's expressed propositional content. We thus have to find some kind of triggering mechanism for the two readings that is not simply up to the communicative intentions of potential speakers in order to maintain this separate pragmatic aspect of the indexical expression. Otherwise, it seems that we do end up with an implicit two-concepts response after all based on a genuine *ambiguity* inherent to the term "actually." For what could such a triggering mechanism be? Stalnaker's two-dimensional semantics allows for one particular type of Gricean triggering which arises based on the violation of either the first or third principle of assertions mentioned above. Stalnaker's framework models a context of conversation wherein diagonal propositions express what we should take an agent to be expressing were they to stand in violation of the principles of assertion while expressing the horizontal proposition. Can the same be said for Comesaña's account of "actually"? Let us see.

<sup>13</sup> As is often accepted, but of course the literature around the border between semantics and pragmatics is vast and multifaceted, so, as everything else in philosophy, it is not a given that these have to be modelled separately. Either way, whether it is a semantic or pragmatic mechanism, it has to be demonstrated.

For the NED problem, since it does not stand in violation of either principle, the diagonal reading will never be triggered. While this is the desired result for the NED problem since it allows acceptance of the conclusion of the NED argument (that the perceptual beliefs of the NED world inhabitants have been produced by actually reliable processes) without contradiction, the diagonal reading cannot be triggered in the context where it *is* desired. This results from the fact that we take rigid reliabilism to be the primary (horizontal) reading. But take again the possible world in which clairvoyance is a reliable process of obtaining true beliefs, which engenders the following argument:

1. A belief is justified if and only if it has been acquired via an actually reliable process.
2. There is a possible world in which clairvoyance is a reliable process of obtaining true beliefs.
3. However, clairvoyance is not reliable in our world and so we need to deny (2) despite of the fact that it is a perfectly reliable process of belief acquisition for the inhabitants of that world.

This argument poses a serious if not insurmountable problem for the “absolutist actualist” (Graham, 2016, 93-94). Now, in order to maintain the feasibility of indexical reliabilism in addition to avoiding the argument above, there needs to be a way to trigger the diagonal reading in this context. That is, the reading where it is indeed true that clairvoyance is an actually reliable process of belief acquisition in the world denoted by the term “actual.” Once again, however, there is (to my understanding) no violation of the principles of assertion, and so we have no reason to think anyone asserting that “S is actually justified in believing that P” (with the actual world and the NED world as the two contexts of utterance and evaluation) would be asserting the diagonal proposition. The proposition is not true in each conversational context under consideration, the assertion is expressing a proposition which has a truth value in each world, and the same proposition is expressed in each world. So, using this strategy, we are left without hope of finding a triggering of the diagonal reading (non-rigid reliabilism) along the lines of Stalnaker’s two-dimensional semantics.

## 5 Conclusion

It is far from obvious how one was to make “actually” behave as an indexical in the regular sense expressed by Kaplan, and even more difficult to see how it would behave in accordance with Gricean maxims when nothing in the context leads us to such considerations. As is described in Graham’s (2016, 94) *Against Actual-World Reliabilism*, it seems that we are in fact dealing with two distinct concepts of actualist reliabilism — indexicalist and absolutist (ibid., 93-94) — both individually untenable:

Surely there are possible, non-actual belief-forming processes (that are as reliable as you please) in possible, non-actual worlds that confer justification on

the beliefs they cause and sustain. Surely there are possible but nonactual epistemically correct belief-forming procedures. Surely the scope of epistemically correct processes transcends the actual processes. Our actual procedures may be correct, but not all correct procedures are actual. This view is clearly mistaken.

Absolutist actual-world reliabilism suffers from the same problem that normal world reliabilism does, according to Graham. It rejects the possibility of non-normal or non-actual reliabilism despite the fact that these phenomena are in no way logically impossible. Actualist reliabilism on its own does not work (and so both varieties, rigid and non-rigid reliabilism, are individually untenable). Only indexicalist actualist-world reliabilism would allow “actually” to function the same way genuine indexicals like “here” and “now” do, that is, serve to make the world of utterance, or some aspect of it, salient (e.g., Diessel 2006, or Burenhult, 2018). However, without an explanation of how the different readings are triggered in a non-arbitrary manner, it seems that we are left clueless when it comes to theorizing with the help of indexicals when they pertain to reliabilism. Lacking an account of the linguistic rules governing the interpretation of “actually,” indexical reliabilism has to be abandoned even though it seems promising in that “actually” can be somewhat accurately modelled with the help of Stalnaker’s two-dimensional semantics in a way that also allows the representation of two notions of reliabilism in one propositional concept that, with the appropriate interchanging of readings, could solve the NED problem and the clairvoyance problem.

Could the reliabilist modify the two actualist concepts within indexical reliabilism into a two-concepts response? This leads to a strange scenario where one could decide freely to apply one of the actual-world reliabilist theories when it avoids a given counterexample, but in each case, one would simultaneously be using an independently indefensible and illegitimate concept in the process. That is to say, absolutist actual-world reliabilism, for instance, solves the NED problem but faces its own counterexamples (the problem of clairvoyance being reliable or other non-normally reliable processes). Indexical actual-world reliabilism avoids the clairvoyance counterexample but cannot handle the NED problem. While one could adopt either of the concepts when it is appropriate to do so (in order to avoid contradiction), as soon as either concept is adopted, the contradictions the chosen concept entails come with it. This scenario would be far more palatable were the indexical aspect of indexical reliabilism satisfyingly understandable, for as a two-concepts response, it is untenable in its own right, and less appealing than other two-concepts responses (e.g., Goldman’s strong and weak justification).

Thus, considering how promising the theory is, I would like to conclude this paper not with some notion of finality or with assumed refutation, but instead to view the discussion above as a challenge for the indexical reliabilist. The field of pragmatics is intricate, and the solution is far from obvious, but it does not seem impossible. And so, the challenge for the indexical reliabilist is to explain how we are to understand the linguistic workings of indexical reliabilism in a way that solves both the NED problem and the clairvoyance problem while maintaining a non-arbitrary linguistic mechanism that determines which of the two readings are at play in the relevant epistemic scenarios.

**Funding** Open access funding provided by Lund University.

## Declarations

**Conflict of Interest** The author declares no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alston, W. P. (1985). Concepts of epistemic justification. *The Monist*, 68, 57–89.
- Alston, W. P. (1993). Epistemic desiderata. *Philosophy and Phenomenological Research*, 53(3), 527–551.
- Armstrong. (2000). The thermometer-model of knowledge. In S. Bernecker & F. I. Dretske (Eds.). *Knowledge: Readings in contemporary epistemology*.
- Ball, B., & Blome-Tillmann, M. (2012). Indexical reliabilism and the new evil Demon. *Erkenntnis*, 78(6), 1317–1336.
- BonJour, L. (1980). Externalist theories of empirical knowledge. *Midwest studies in philosophy*, 5, 53–73.
- Burenhult, N. (2018). 17 The Jahai multi-term demonstrative system: What's spatial about it? *Demonstratives in cross-linguistic perspective*, 14, 361.
- Cohen, S. (1984). Justification and truth. *Philosophical Studies*, 46(3), 279–295.
- Comesaña, J. (2002). The diagonal and the demon. *Philosophical Studies*, 110(3), 249–266.
- Comesaña, J. (2010). Evidentialist reliabilism. *Noûs*, 44(4), 571–600.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17, 463–489.
- Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press.
- Foley, R. (1985). What's wrong with reliabilism? *The Monist*, 68, 188–202.
- Goldman, A. I. (1988). *Strong and weak justification. Philosophical perspectives*, 2, 51–69.
- Graham, P. J. (2016). *Against actual-world reliabilism. : Foundations and applications*. Oxford University Press
- Grice, P. (1989). *Studies in the way of words*. Harvard University Press.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), (1989). *Themes from Kaplan*. Oxford University Press.
- Kornblith, H. (1983). Justified belief and epistemically responsible action. *The Philosophical Review*, 92(1), 33–48.
- Kripke, S. (1981). Naming and necessity (Rev. and enl. ed.).
- Lehrer, K., & Cohen, S. (1983). Justification, truth, and coherence. *Synthese*, 55(2), 191–207.
- Lewis, D. (1970). Anselm and actuality. *Noûs*, 4, 175–188.
- Moon, A. (2012). Three forms of internalism and the new evil demon problem. *Episteme*, 9(4), 345–360.
- Pollock, J. (1984). Reliability and justified belief. *Canadian Journal of Philosophy*, 14, 103–114.
- Pryor, J. (2001). Highlights of recent epistemology. *The British Journal for the Philosophy of Science*, 52(1), 95–124.
- Sosa, E. (1993). Review: Proper functionalism and virtue epistemology. *Noûs*, 27(1), 51–65.
- Sosa, E. (2001). Goldman's reliabilism and virtue epistemology. *Philosophical Topics*, 29(1/2), 383–400.
- Sosa, E. (2003). *Epistemic Justification: Internalism vs. externalism, foundations vs. virtues* Blackwell Publishing.
- Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics*. Academic Press.
- Stalnaker, R. C. (1999). *Context and content: Essays on intentionality in speech and thought*. Clarendon Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.