A NEW RESPONSE TO THE NEW EVIL DEMON PROBLEM

Umut BAYSAN

ABSTRACT: The New Evil Demon Problem is meant to show that reliabilism about epistemic justification is incompatible with the intuitive idea that the external-world beliefs of a subject who is the victim of a Cartesian demon could be epistemically justified. Here, I present a new argument that such beliefs can be justified on reliabilism. Whereas others have argued for this conclusion by making some alterations in the formulation of reliabilism, I argue that, as far as the said problem is concerned, such alterations are redundant. No reliabilist should fear the demon.

KEYWORDS: dispositions, justification, The New Evil Demon Problem, reliabilism

The New Evil Demon Problem, presented by Cohen,[1] is meant to show that reliabilism of the sort that was defended by Goldman[2] is incompatible with the intuitive idea that the external-world beliefs of a subject who is the victim of a Cartesian demon *could* be epistemically justified. The original argument goes as follows:

(1) If reliabilism is true, no external-world belief of a victim of an evil demon could be justified.

(2) Some external-world beliefs of the victims of an evil demon could be justified.

(3) Therefore, reliabilism is false.

One might think that there can't be much to add to the debate over the New Evil Demon Problem after more than thirty years of discussion. Nevertheless, there remains a *prima facie* plausible solution to this problem which hasn't been quite stated. In what follows, I shall present this solution.

As formulated as an objection to reliabilism, the argument that is sketched above takes reliabilism to be the view that a belief is justified if and only if it is formed as a result of a reliable belief-forming process. I shall call this version of reliabilism *crude reliabilism*. Just to restate, (3) holds that crude reliabilism is false.

[1] Stewart Cohen, "Justification and Truth," *Philosophical Studies* 46 (1984): 279-296.

[2] Alvin Goldman, "What Is Justified Belief?" in *Justification and Knowledge,* ed. G. Pappas (Ithaca: Cornell University Press, 1979), 1-23.

The reason for thinking that (1) is true is the following. Most (if not all) external-world beliefs of a victim of an evil demon are false because in demon worlds radical sceptical hypotheses are true: in a demon world, either there is no external world or the external world is radically different from the way it appears. Given the high frequency of false beliefs, the belief-forming processes of the habitants of demon worlds cannot be reliable; hence their beliefs cannot be epistemically justified. Or so the objector thinks.

The intuition that supports (2), which I shall call *the fairness intuition*, is that in some cases, victims of an evil demon might be doing all the right things in order to hold true external-world beliefs. When they believe that there are trees and cats in their surroundings, they do so because they undergo perceptual experiences that are subjectively indistinguishable from those experiences which are truly caused by real trees and cats. In fact, a victim of an evil demon could be an *epistemic counterpart* of you, and we might suppose that you have mostly epistemically justified beliefs. Here, I take an epistemic counterpart of S at t_1 to be someone whose beliefs are, as far as their narrow contents are concerned, type-identical with the beliefs of S at t_1, and are furthermore held for the same subjectively accessible reasons as those of S at t_1. To illustrate: you believe at 11am today that it will rain; your reason for holding this belief is that you have a memory of the weather forecast reporting that it would rain. Your epistemic counterpart (as far as your temporal part at 11am today is concerned) believes that it will rain, and her reason for holding this belief is that she has a memory of the weather forecast reporting that it would rain. And it may be the case that whereas your belief is true, your epistemic counterpart's belief is false (or *vice versa*). The fairness intuition, I think correctly, suggests that if your beliefs are mostly justified, then your epistemic counterparts' beliefs should be mostly justified too. Assuming that you and your epistemic counterpart have the same reasons for holding same beliefs – you had the same perceptual experiences, have used the same inference rules, and so on – it is only *fair* to expect that your beliefs are justified if and only if your epistemic counterparts' beliefs are justified.

Let me just briefly state what I will *not* argue for. I will *not* argue that crude reliabilism can be weakened, or relativised, or made indexical, in order to accommodate the fairness intuition. Strategies along those lines have been endorsed by others,[3] and they do have their own virtues. But I believe that it is worth noting that such routes are *redundant*, at least as far as the New Evil Demon

[3] Alvin Goldman, *Epistemology of Cognition* (Cambridge: Harvard University Press, 1986) and Juan Comesaña, "The Diagonal and the Demon," *Philosophical Studies* 110 (2002): 249-266.

Problem is concerned. Crude reliabilism, without any further qualification, can accommodate the fairness intuition; we can formulate epistemic justification as the reliability of belief-forming processes, and still hold that our demonic epistemic counterparts' beliefs can be justified.

The key is to recognise that 'reliable' is a *dispositional* concept and, arguably, *reliability* is a dispositional property – insofar as it is a real property and there are dispositional properties. If one has problems with the idea of dispositional properties, most of what I will say can be understood in a non-dispositionalist framework. Take a true dispositional expression: "This vase is fragile." Why is this statement true? A full-blown realist about dispositional properties would say it is true *because* the vase that the "the vase" refers to is a bearer of the dispositional property of *being fragile*. Someone who is sceptical about dispositional properties, however, would say that the truth of this expression consists in the fact that the vase in question has some non-dispositional properties such that having those properties in the right circumstances makes it the case that the vase behaves in a fragile manner. The upshot is this: one needn't be a full-blown realist about dispositional properties in order to make sense of dispositional expressions.

Now consider *reliability* as a dispositional property. Take a supposedly true dispositional expression, such as "Lily is reliable." Whereas a realist about dispositionalist properties would say that this expression is true in virtue of the fact that Lily instantiates a dispositional property, namely *reliability*, an anti-realist about dispositional properties can still give a non-dispositional truthmaker about Lily for the said expression. I don't really want to be committed to any view about the reality or fundamentality of dispositional properties, but the points that I will make are easier to express with the resources of a dispositionalist view, so I will treat *reliability* as a dispositional property.

Many *sorts* of things can be reliable, and likewise, unreliable: people, machines, newspapers, weather, Wi-Fi signals, belief-forming processes, so on and so forth. When I say that Lily is reliable, arguably, I am not referring to the very same property of *reliability* that I refer to when I say that the Wi-Fi signal is reliable. A person's reliability consists in her disposition to tell the truth (or what she takes to be the truth) and keep her promises in the right circumstances. When the circumstances are not right, however, a reliable person might be forced to lie, or break a promise.

A belief-forming process's reliability consists in something quite different. Whereas the reliability of Lily is manifested in her telling the truth in the right circumstances, the reliability of a belief-forming process is manifested in the fact

that beliefs that are formed as a result of that process are mostly true, again, in the right circumstances. A reliable belief-forming process is disposed to produce true beliefs. That is, the manifestation of the dispositional property *reliability* attributed to a belief-forming process is the *truth* of the belief that is formed. Unreliable belief-forming processes, such as wishful thinking, aren't disposed to produce true beliefs. Occasionally, the beliefs that are formed as a result of wishful thinking may turn out to be true. But this is not different from the occasional breaking of non-fragile vases. Such occasional breakings don't have to be miraculous. Vehicles like the Popemobile and the Batmobile have windows made of non-fragile glass, yet presumably they couldn't stay intact after an atomic bomb explosion. Our standards for non-fragility are not so high that only absolutely unbreakable things can be deemed non-fragile.

Although the reliability of a person and the reliability of a belief-forming process might be different properties, the rules of the application of the predicate 'is reliable' to people and to belief-forming processes are similar in an interesting way. The similarity lies in the fact that *one can be a bearer of a dispositional property without ever manifesting the disposition in question*. Strictly speaking, it is possible for a reliable person to lie *at all times*. Admittedly, this sounds very odd; nevertheless it is true. It belongs to the concept of 'disposition' that dispositions needn't be manifested in order to be instantiated. There are fragile vases which are never broken, simply because they have never been struck. So, the following is a perfectly possible state of affairs:

(i) *a* is fragile; *a* is not struck; *a* doesn't break.

But more strangely, there *could be* fragile vases that are never broken, despite being struck and dropped multiple times. Think of the case of the sorcerer who is the guardian of a fragile vase.[4] Every time the vase is struck, the sorcerer casts a spell on it so that it resists the strike. The vase in question still counts as fragile; if the sorcerer weren't guarding it, it would have manifested its fragility. (Note that this is true non-vacuously: the sorcerer is only contingently protecting the vase.) So, the following is a perfectly possible state of affairs too:

(ii) *a* is fragile; *a* is struck many times; *a* doesn't break.

Moving on from fragile vases to reliable people: consider the case of Lily. Lily is disposed to tell the truth, but for some reason, at every single attempt, she fails to do so. Maybe her actions are being manipulated by the Purple Man, who is a master of mind-control. Lily wants to tell the truth; she genuinely intends to do

[4] David Lewis, "Finkish Dispositions," *Philosophical Quarterly* 47 (1997): 143-158.

so, but every time she speaks, she lies. She, I stipulate, is still reliable, but she is not manifesting her reliability, because she is being controlled by the Purple Man. If the Purple Man weren't manipulating her actions, Lily would have told the truth. (Again, this is true non-vacuously: the Purple Man is only contingently manipulating Lily's decisions.) So, the following is a possible state of affairs:

> (iii) *a* is a reliable person; *a* is asked if P is true; *a* knows that P is true; *a* says that P is false; this happens systematically.

I hope I have convinced you that (ii) and (iii) are possible states of affairs. If you still have doubts, remember that dispositions require right circumstances in order to be manifested in the right way. By introducing sorcerers and mind-controlling supervillains, we are departing from right circumstances.

Now, beliefs. A belief forming-process may be disposed to produce true beliefs, but for whatever reason, at every attempt, it may fail to do so. As I hope is clear from the discussion so far, all we need to do is depart from the right circumstances. In a demon world, what is happening is exactly this. The deeds of the evil demon change the circumstances so the belief-forming processes, however reliable they are, are not manifesting their reliabilities. So, the following is also a perfectly possible state of affairs:

> (iv) *a* is a reliable belief-forming process; *a* is exercised; *a* doesn't produce true beliefs; this happens systematically.

Now if (iv) is really a possible states of affairs, premise (1) of the argument above is false: one can be a crude reliabilist about epistemic justification and still hold that external-world beliefs in a demon world can be epistemically justified. If all this is right, then it appears that crude reliabilism doesn't have to be weakened or relativised in order to accommodate the fairness intuition. What needs to be done is to recognise that *reliability* is a dispositional property and remember that dispositions can be held without ever being manifested.

If I am right, crude reliabilism, a version of reliabilism which has been abandoned partly due to worries about the New Evil Demon Problem, actually has the resources to deal with this problem. I showed a hitherto unexplored and *prima facie* plausible logical space where both crude reliabilism and the fairness intuition are true.[5]

[5] **Acknowledgments.** Many thanks to Robert Cowan and Martin Smith for discussion and comments on a previous version of the paper. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.