JUAN COMESAÑA

THE DIAGONAL AND THE DEMON

(Received in revised form 27 June 2002)

ABSTRACT. Reliabilism about epistemic justification – the thesis that what makes a belief epistemically justified is that it was produced by a reliable process of belief-formation – must face two problems. First, what has been called "the new evil demon problem", which arises from the idea that the beliefs of victims of an evil demon are as justified as our own beliefs, although they are not – the objector claims – reliably produced. And second, the problem of diagnosing why skepticism is so appealing despite being false. I present a special version of reliabilism, "indexical reliabilism", based on two-dimensional semantics, and show how it can solve both problems.

1. INTRODUCTION

What makes a belief epistemically justified? Reliabilists about justification have a nice answer: what makes a belief epistemically justified is that it was produced by a reliable process of belief-formation – i.e., a process most of whose outputs are true (or, more precisely, a process which, when used under the appropriate circumstances, would yield mostly true beliefs as outputs):

> **Reliabilism**: a belief is epistemically justified if and only if it was produced by a process most of whose outputs would be true.[1]

Stewart Cohen (1984) presented a counterexample to reliabilism about justification. Cohen's objection has been called "the new evil demon problem",[2] and he stated it thus:

> Imagine that, unbeknown to us, our cognitive processes (e.g., perception, memory inference) are not reliable owing to the machinations of the malevolent demon. It follows on a Reliabilist view that beliefs generated by those processes are *never* justified. (. . .) Is this a tenable result? I maintain that it is not. (. . .) [P]art of what the hypothesis entails is that our experience is just as it would be if our cognitive processes were reliable. Thus, on the demon hypothesis, we would have every reason for holding our beliefs that we have in the actual world. (p. 281)[3]

Philosophical Studies **110:** 249–266, 2002.

The new evil demon problem arises from:

> **The new evil demon thesis**: the beliefs of the victims of evil demons (henceforth called "demoners") are as justified as our own beliefs.

Cohen's claim is that reliabilism is incompatible with the conjunction of the new evil demon thesis and some obvious truths.

The new evil demon thesis can also be put to another use. The skeptic and the Moorean agree that if we know that here is a hand, then we also know that we are not demoners. But the skeptic and the Moorean disagree on whether to apply Modus Ponens and conclude that we do know that we are not demoners, or Modus Tollens and conclude that we don't, after all, know that here is a hand. Clearly, the Moorean answer is, at least *prima facie*, the right one. But the skeptic might try to get beyond the *prima facie* by arguing in the following way:

> **Tempting but fallacious argument (TFA)**: Surely we are justified in believing that we are not demoners: everything tells against that possibility and nothing tells for it. But, given the new evil demon thesis, we would be equally justified even if we *were* demoners. But, of course, if we were demoners we wouldn't know that we are not demoners. Therefore, the justification that we have for believing that we are not demoners is not the justification necessary to make that belief count as knowledge.

The new evil demon problem, therefore, is doubly problematical for a reliabilist Moorean: it is a direct problem for her reliabilism, and an indirect problem for her Mooreanism. What is a reliabilist Moorean to do? She can, of course, reject the new evil demon thesis. But there is another, less bullet-biting alternative: Ernest Sosa has suggested a version of reliabilism, what I will call "indexical reliabilism", which is *not* incompatible with the new evil demon thesis.[4]

In section two of this paper I will present indexical reliabilism using a two-dimensional semantics framework (which allows us to see clearly the connection between the theory and the new evil demon problem). In section three, I will show how indexical reliabilism allows us to diagnose a fallacy in TFA, and I will lay out some consequences of indexical reliabilism having to do with the relations between epistemic justification and warrant, understood as whatever has to be added to true belief in order to get knowledge.

2. INDEXICAL RELIABILISM

2.1. *Two-dimensional Semantics*

I enter the room and I say “Michael is here”. In the actual world (W0), I am entering the lounge in the Philosophy Department, and Michael is indeed there. There is a possible world (W1) where I am entering the same room but Michael is in the cafeteria, and there is another possible world (W2) where I am entering the cafeteria, but Michael is in the lounge. The proposition that I expressed is that Michael is in the lounge, and so it is a proposition that is true in W0 and W2 but false in W1, where Michael is in the cafeteria. Never mind that, in W2, *I* am in the cafeteria, and so I say something false in W2 (that is, I would have said something false, had I been in the cafeteria). I, in fact, am *not* in the cafeteria, and so what I in fact said is true in W2. If propositions are functions from possible worlds to truth-values, and if the only relevant worlds are the ones mentioned so far, the proposition that I expressed by saying “Michael is here” can be represented by the following matrix:

A

W0	W1	W2
T	F	T

What I say (what proposition is expressed by the sentences I utter, for example) depends on how the actual world is. This is, in a sense, trivially true, for how the actual world is includes what sentences I utter, and also what language I use. But there is a subtler way in which the actual world determines what I say *when I use indexical terms*. If we consider, for example, W2, the proposition expressed by my uttering “Michael is here” in that world is the proposition that Michael is in the cafeteria, a proposition that is true in W1 but false in W0 and W2. That proposition is represented by the following matrix:

B

W0	W1	W2
F	T	F

The proposition that I express at W1 (that is, the proposition that I would have expressed had Michael been in the cafeteria while I was entering the Lounge) is the same as the one that I actually expressed. Michael's location is the same in W0 as in W2, but what I say is the same in W1 as in W0. To represent all this information, we can use the following two-dimensional matrix (where each row represents the proposition expressed when considering the world on the left as actual):[5]

C

	W0	W1	W2
W0	T	F	T
W1	T	F	T
W2	F	T	F

Two-dimensional matrices are helpful not only in representing the way the context determines the content of our utterances: they also provide a solution to a puzzle about communication. Suppose that Dave has a meeting at 3:30 and doesn't know what time it is now, and so he asks me. I reply "It is now three o'clock". Now, assuming that the conversationally relevant possible worlds (those among which Dave cannot distinguish) are a world where it is now three o'clock, a world where it is now three-thirty, and a world where it is now four o'clock, it would seem that I haven't given Dave any useful information. Given that 'now' is an indexical (an utterance of 'now' at time *t* refers to *t*), what I said is that three equals three. But that proposition is necessarily true. Moreover, had I been wrong, I would have expressed a necessarily false proposition. What I said, together with what I could have said, can be represented in the following matrix (where '3:00' is the name of the possible world where it is now three o'clock, etc.):

D

	3:00	3:30	4:00
3:00	T	T	T
3:30	F	F	F
4:00	F	F	F

Now, given that Dave is not sure which of 3:00, 3:30 or 4:00 is the actual world, it seems that the information contained in the above matrix won't help him, *for no row of the matrix distinguishes among the three worlds Dave is unsure about*. Again, depending on which world we are in, I expressed a proposition that is either necessarily true or necessarily false. Now, Dave, of course, is no fool, and he correctly understands me as meaning that 3:00 is the actual world.[6] Somehow, by saying "It is now three o'clock", I managed to transmit to Dave the information that 3:30 and 4:00 should be thrown away as relevant possible worlds in the conversational context. One way of formalizing Dave's implicit interpretational strategy is to say that he applies a two-dimensional operator to D: the operator which, in Stalnaker's words, "takes the diagonal proposition and projects it onto the horizontal".[7] The diagonal proposition of a given matrix is the one whose values can be read along the diagonal of the matrix. In this case, the application of the operator to D gives us:

E

	3:00	3:30	4:00
3:00	T	F	F
3:30	T	F	F
4:00	T	F	F

This last matrix is now *obviously* helpful to Dave, for it is transmitting the same information no matter which world is actual, and the information is (how else would you put it?) that it is now three o'clock.

When Dave "diagonalizes" the matrix that represents my utterance, this is not something that he does consciously – indeed, there is *nothing* that he does that is different from what he does in interpreting any of my utterances, whether they contain indexicals or not. Rather, Stalnaker's two-dimensional operator and his diagonal proposition are a graphically vivid way of representing an interesting feature of indexical terms.[8] The content of an indexical term is determined by the context of use together with a function from contexts to contents – a function that Kaplan called the "character" of an indexical.[9] This feature of indexicals allows for two kinds of information that we might want to transmit (or that we might be interpreted as transmitting) when uttering sentences with indexical terms. When I say to you "Michael is here", for example, I might want to transmit the information contained in the proposition represented by matrix A above. But I might also want to transmit a different kind of information. Suppose that you ask me "Are we in the lounge or in the cafeteria?", and I answer "I don't know, but Michael is here". In that case, although I expressed the proposition represented by A, that is not the information that I wanted to transmit, and that is not the proposition that you grasped when you interpreted my utterance. Rather, the information that was successfully transmitted in our exchange is the proposition that is true in W0 and false in W1 and W2 – in other words, the diagonal proposition of matrix C.

2.2. *Reliability*

Let us restrict our discussion in the following ways. We will only talk about the epistemic status of *perceptual* beliefs. We will assume that how the world is in fact, how it seems to epistemic agents in experience, and how epistemic agents react doxastically to how the world seems to them in experience, are three independent variables. We take the actual world to be one in which, nearly enough, when it is the fact that *p*, the world presents itself in our experience as if *p*; and, as it happens, we nearly enough react with the belief that *p* whenever the world presents itself in our experience as if *p*. But we can conceive of worlds where there is no such correlation between how things are, how they seem to be, and the doxastic reaction of the epistemic agents (demoners inhabit such a world). The assumption of *total* independence may be too strong to be true, but it is harmless in the present context, for the rest of the discussion could be carried out under less drastic restrictions. On the other hand, *some* independence between the three elements is required for the new evil demon problem to even arise. If, as some externalists about the content of experience might want to hold, the content of experience is *determined* by how the world is in fact, so that, for example, an experience represents whatever it is reliably connected to in the external world, then the experiences of demoners would *not* be like ours.[10]

Let us now spell out the new evil demon problem in somewhat greater detail. Let us assume, for reductio, that reliabilism is correct:

1. **Reliabilism**: a belief is epistemically justified if and only if it was produced by a process most of whose outputs would be true.

Now we also assume a mildly anti-skeptical thesis:

2. Our beliefs are justified.

The crucial premise is, of course, the new evil demon thesis:

3. **The new evil demon thesis**: the beliefs of demoners are as justified as our own beliefs.

From 2 and 3 we can infer:

4. The beliefs of demoners are justified.

And 1 (in its left-to-right direction) and 4 now give us:

5. **Demonic reliability**: the beliefs of demoners were produced by a process most of whose outputs would be true.

But, Cohen claims, **Demonic reliability** is clearly false. Therefore, reliabilism is false.

What indexical reliabilism denies is the last step in the reductio. There is a clear sense in which **Demonic reliability** is true, and the sense in which it is clearly false is one that the indexical reliabilist can account for. But let us first take a look at the other possibilities left open by Cohen's argument. We could deny 2, and some skeptics might even take this argument as favoring their position. A straight denial of 2, however, seems an overreaction. Later, we shall see that the skeptic has subtler (though ultimately unsuccessful) ways of using Cohen's argument to his own advantage. We could also deny the central premise 3, and this has been the favorite option of some reliabilists. But Cohen's appeal to the new evil demon thesis is an appeal to intuition, and given that 3 does capture an intuition that many have, to reject it would put the reliabilist at a dialectical disadvantage. Indexical reliabilism, we shall see, can explain the appeal of 3 while, at the same time, accounting for the discomfort that some reliabilists feel towards it.

How could **Demonic Reliability** be true? Isn't it just obvious that the demoners' beliefs are produced by *unreliable* methods? Let us say that the process by which they acquired their beliefs can be described as "taking perception at face value", though nothing hangs on the exact choice of description.[11] In the demoners' world, taking experience at face value is highly *un*reliable: it usually yields false beliefs, and it would yield mostly false beliefs in counterfactual applications – *in the demoners' world*. Here, in this world, taking experience at face value is highly reliable: it usually yields true beliefs, and this is not just a statistical correlation, for it *would* yield true beliefs if used in appropriate circumstances. So the crucial question is what we are referring to when we say that a justified belief must have been produced by a process *most of whose outputs would be true*. True where? The immediate answer is: true *simpliciter*, that is, true in the actual world. But, under that understanding of reliabilism,

> **Reliabilism 1**: a belief is epistemically justified if and only if it was produced by a process that is actually reliable,

we can see that Cohen's argument from 1 to 5 can go through as before, but now 5 is no longer clearly false. Indeed, in a straightforward interpretation, it is clearly true, as clearly true as the fact that taking experience at face value is reliable here.

Now, as David Lewis has argued, "actual" itself is an indexical term: for any world *w*, an utterance of "actual" in *w* refers to *w*.[12] Suppose now that I say "Ernie's beliefs are justified", where the meaning of "justified" is given by **Reliabilism 1**. Assuming that the relevant worlds are the actual world (W0) and a world where everybody is a demoner (W3), the following matrix represents what I said and what I could have said:

F

	W0	W3
W0	T	T
W3	F	F

Given how the actual world is in fact, what I said is true both in the actual world and in W3. Had I been a demoner, on the other hand, I would have said something different, something that is false both in W0 and W3. (Caution: it is clear that I would have expressed a different proposition, but would I have *said* something different? There is room for disagreement here. Clearly, I would have uttered the same sentence, and spoken the same language. I would have said something different only in the sense in which I say something different when I say "I wish you were here", speaking in different places to different persons.[13]) In other words, the sentence "Ernie's beliefs are justified" could have expressed a false proposition, had it been uttered in a demoner world. But, given how the actual world is in fact, Ernie's beliefs are justified, and they would still have been justified had he been a demoner. So **Demonic Reliability** is true, although the sentence that expresses this thesis could have expressed a false proposition.

Now, as is also noted by Lewis, there is "a complication" in the indexical analysis of actuality:

> [W]e can distinguish primary and secondary senses of "actual" by asking what world "actual" refers to at a world w in a context where some other world v is under consideration. In the primary sense, it still refers to w, as in "If Max ate less, he would be thinner than he actually is". In the secondary sense it shifts its reference to the world v under consideration, as in "If Max ate less, he would actually enjoy himself more".[14]

Let W0 be, again, the actual world. When you are considering an arbitrary world V and you say that the beliefs of its inhabitants are justified, you are saying, according to the indexical reliabilist, that they were produced by processes that are reliable in the actual world. But you could be using "actual" in its primary or in its secondary sense. If V is W0, there is no real difference, of course; but if V is different from W0, then you could be saying that their beliefs were produced by processes that are reliable in W0, or you could be saying that their beliefs were produced by processes that are reliable in V. In other words, when attributing justification to the beliefs of merely possible epistemic agents, you might be saying that they are produced by processes that are reliable here, or you might be saying that they are produced by processes that are reliable in their world. This last thing that you might be saying is captured by this definition of justification:

> **Reliabilism 2**: 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.

Taking into account the “complication” in the behavior of “actual” in counterfactual contexts noted by Lewis, we should not take **Reliabilism 1** and **Reliabilism 2** as characterizing two independent notions. Rather, we should think that a notion with the application conditions of the right-hand side of **Reliabilism 2** is expressed, in some contexts, by using the notion characterized in the right-hand side of **Reliabilism 1** – just as we sometimes use “now” in its diagonalized sense, though we don’t think that we have two notions of “now”.[15] That, indeed, is what indexical reliabilism amounts to.

We are now in a position to see that there is a sense in which **Demonic Reliability** is false (false in the actual world, false *simpliciter*). If what I (want to) express, when I say “Ernie’s beliefs are justified”, is the diagonal proposition in F, namely:

G

	W0	W3
W0	T	F
W3	T	F

then I am saying that Ernie’s beliefs are justified, though they wouldn’t have been so justified if he had been a demoner. Another way of expressing the proposition represented by G is by saying that the demoners’ beliefs are not justified (while ours are) in the sense of justification characterized by **Reliabilism 2**. “Ernie’s beliefs are justified”, then, is ambiguous; and so is **Demonic Reliability**, and for the same reasons. (I am stretching a bit the notion of ambiguity to put my point succinctly. I don’t mean to suggest that “reliable process” has two meanings, just as I wouldn’t suggest that “now” has two meanings.) If interpreted according to **Reliabilism 1** it is true but it poses no danger to reliabilism, for there is no intuition that it should be false. If interpreted according to **Reliabilism 2**, on the other hand, it is false. But **The new evil demon thesis** is also false when justification is understood according to **Reliabilism 2**, as we shall see in section 3.[16] But, before discussing that, there is one worry that we should address.

2.3. *Time and Space*

So far, indexical reliabilism is able to deal with the original version of the new evil demon problem, where the demoners are located in another world. But there is an analogous problem for reliabilism if we locate the demoners not in a different world but at a different time. Suppose, then, that a malevolent demon is waiting in limbo for the year 3002 to come, when he will victimize everyone. In that case, although the beliefs of the subjects of 3002 are unreliably produced, they are intuitively just as justified as ours. But, given that they are located in the future in the actual world, indexical reliabilism doesn't have the resources necessary to deal with them.[17]

The solution is entirely analogous to the original case. The beliefs of the subjects in 3002 are horizontally justified because the processes by which their beliefs are produced are actually reliable *now* (to be redundant, in 2002). Their beliefs still lack something that ours have, though, and that is having been produced by processes that are actually reliable at the time when they are used. Formally, the solution consists in taking pairs of worlds and times (as opposed to only worlds) as the two-dimensional indices in the evaluation of the truth-value of sentences involving "justified".

But there still seems to be a new evil demon problem that indexical reliabilism is unable to deal with. What if the demoners are located, not in another world, not even at another time, but in another place, say, in a faraway galaxy? Are the processes by which their beliefs are produced actually reliable now or not? (And what about our processes? Aren't they the same as theirs?) Well, they are reliable *here*, but not *there*. Their beliefs are horizontally justified because the processes by which they are formed are actually reliably *here* and now. Their beliefs still lack something that ours have, though, and that is having been produced by processes that are actually reliable at the time *and place* where they are produced. Formally, the solution consists in taking centered worlds (triples of worlds, times and places) as the two-dimensional indices in the evaluation of the truth-value of sentences involving "justified".[18]

3. SKEPTICISM, JUSTIFICATION, AND WARRANT

3.1. *The Rebuttal of the Tempting but Fallacious Argument*

We can restate TFA thus:

(A) The demoners' belief that they are not demoners lacks something by way of justification, and this lack is in part responsible for the belief's not qualifying as knowledge.
(B) Our belief that we are not demoners doesn't have more, by way of justification, than the demoners' belief. Therefore,
(C) Our belief that we are not demoners doesn't qualify as knowledge.

Armed with indexical reliabilism, we can now clearly see why TFA is *both* tempting and fallacious. It is fallacious because it is a fallacy of equivocation. It is tempting because it has true premises and it is a particularly subtle fallacy of equivocation.

There are two interpretations of (A) and two interpretations of (B), according to whether we are attributing the horizontal or the diagonal sense of reliability. Those interpretations yield the following propositions (where P1 is the process that produced the demoners' belief that they are not demoners and P2 is the process that created our belief that we are not demoners):

(A1) P1 is not reliable in W0, and that is part of the explanation of why demoners lack knowledge.
(A2) P1 is not reliable in W3, and that is part of the explanation of why demoners lack knowledge.
(B1) The reliability of P2 in W0 is at most as great as the reliability of P1 in W3.
(B2) The reliability of P2 in W0 is at most as great as the reliability of P1 in W0.

The argument from (A1) and (B2) to (C) is valid, as is the argument from (A2) and (B1) to (C), but both (A1) and (B1) are false. On the other hand, (A2) and (B2) are both true, but they don't yield (C).[19]

3.2. *Justification and Warrant*

We explained away the appeal of the skeptical argument by pointing to a certain kind of context relativity in the notion of justification. But many have been tempted to reply to the argument by saying that *warrant*, understood as whatever has to be added to true belief to yield knowledge, is radically disjoint from justification. If that were so, there would be no mystery as to how demoners could be as justified as we are and yet lack knowledge. With that answer to the skeptical argument, however, comes a certain disdain for the notion of knowledge. The paradoxical idea seems to be that warrant (and, therefore, knowledge) is not epistemically relevant, precisely *because* warrant is disjoint from justification.[20] Indexical reliabilism can help us avoid that conclusion.

Note that the notion characterized in *Reliabilism 2* is necessary for knowledge. What the demoners' beliefs lack is the kind of justification that accrues to a belief by having been produced by a process that is reliable in the world where the belief is held. Justification in the sense of diagonal reliability (that is, reliability in the world that the belief inhabits) is necessary for warrant, whereas justification in the sense of horizontal reliability (reliability here) is not.[21] In other words, "Reliability is necessary for warrant" is true if by "reliability" we mean diagonal reliability, but false if we mean horizontal reliability. It follows that there can be (horizontally) justified beliefs that are not warranted (as already shown by the case of demoners), and that there can be warranted beliefs that are not (horizontally) justified (as shown by the possible world where information can only be acquired by processes that are unreliable here). This is a way in which justification and warrant can come apart – but they can *thus* come apart only for non-actual epistemic agents. Given that to say that our beliefs are reliably produced in the world where they are held is the same as saying that they are reliably produced in the actual world, our beliefs (that is, the beliefs of actual epistemic agents) have the kind of justification necessary for warrant if and only if they have the other kind of justification.

Indexical reliabilism, then, both suggests a strong connection between epistemic justification and warrant and shows a way in which they can come apart. The connection is not, alas, an implicational one: warrant is not necessary for (horizontal) justification, nor is (horizontal) justification necessary for warrant. What indexical reliabilism shows is that, despite the independence of (horizontal) justification and warrant noticed in the previous paragraph, we don't need two radically different notions, one to characterize epistemic justification and the other to characterize warrant. One indexical notion can do both jobs.[22] This, in turn, suggests that the internalist consequences that some have drawn from the new evil demon problem are uncalled for.

Indexical reliabilism, then, allows that a belief can be warranted (and thus, if true, amount to knowledge) even if it is not (horizontally) justified. Doesn't this present a problem for us? Consider a world of wishful thinkers aided by a benevolent demon that makes wishful thinking reliable. Their beliefs are clearly not (horizontally) justified. But aren't they warranted, given that they are diagonally reliable? Don't they then have knowledge? And isn't this counterintuitive?

Let me start with that last question. What is our intuition regarding a case of wishful thinkers that are epistemically successful owing to the interventions of a benevolent demon? Do they know or don't they? To answer that question it would be important to know what precisely it is that the benevolent demon does. One possibility is that the benevolent demon creates a world where wishful thinking is reliable, then creates wishful thinkers, and then just contemplates his creation. Another one is that the benevolent demon interferes with a world of wishful thinkers so as to make their beliefs reliable – but the world is such that, were it not for the demon's constant intervention, it would make wishful thinking go wrong most of the time. My intuition is that wishful thinking does provide knowledge in the first case (although it might be that "wishful thinking" is a misnomer in that case) but not in the second. (I see the first case as similar to the case of subjects who inhabit a world where light travels in funny ways, so that, e.g., when they "see" something in front of them, it is really to one side, but are so constituted that when they "see" something in front of them they form the belief that it is to one side.)

But I agree that, in the second case, wishful thinking doesn't provide knowledge – but how can I say *that*, given that their beliefs are diagonally reliable? I can say it because diagonal reliability is only a necessary condition for warrant, not a sufficient one. Now, there is still the question, “Why is it that they lack knowledge, if not because their beliefs are not diagonally reliable?” To answer that question would involve saying what more, besides diagonal reliability, is necessary for warrant – and I really don’t know that. Now, I have a suspicion, although I will not argue for it right now. My suspicion is that, in that case, what’s wrong with wishful thinkers is that the reliability of their beliefs doesn’t have to do with how well their cognitive processes are adjusted to their environment, but is rather a matter of chance. And this also explains my intuition that in the first sort of case wishful thinking does provide knowledge. But I realize that to make this a substantial answer I would have to give an account of the difference between good adjustment to your environment and adjustment just by chance, and that is what I am not prepared to do.

NOTES

[*] Earlier versions of this paper were presented at the 2001 Graduate Conference in Epistemology, Rochester University, the 2001 Harvard-MIT Graduate Student Philosophy Conference, the 2001 Columbia Graduate Student Philosophy Conference, and at the XIX International Philosophy Congress, Skepticism, Instituto de Investigaciones Filosóficas, UNAM, México; I am grateful to the audiences at those conferences for their questions and comments. Many thanks to Stewart Cohen, James Dreier, Jordi Fernández, Mark Heller, David Matheson, Agustín Rayo, Joseph Shieber and Gabriel Uzquiano for very helpful comments and criticisms. Special thanks to Manuel Comesaña, Michael Pace, Carolina Sartorio, and Ernest Sosa, who read several drafts and offered extremely generous doses of suggestions and support. Research leading to this paper was supported in part by a grant from Fundación Antorchas.

[1] The counterfactual condition is needed in order to account for the possibility of highly reliable but never used processes, on the one hand, and of processes that actually yield lots of true beliefs but are intuitively unreliable because they would very easily yield false beliefs. It would perhaps be closer to the truth to hold that reliability is a necessary, though not sufficient, condition for epistemic justification, and much more is needed in order to fully characterize an epistemically useful notion of reliability. But the problems and issues that will be examined here can be dealt with at this rough level.

[2] I am not sure who baptized the problem. I learned it from Ernest Sosa (1991).

[3] In the scenario that Cohen is imagining, are our beliefs true or false? Cohen doesn’t say, but both possibilities represent counterexamples to reliabilism about

justification – in the scenario where our beliefs are still true, the demon is tampering with our belief-forming processes so as to render them counterfactually unreliable, i.e., such that they would very easily yield mostly false beliefs. In the paper, I don't distinguish between these two possible skeptical hypotheses except where it matters. One might think that the case where our beliefs are still true is like a generalized Gettier case – but it is not sufficiently like a Gettier case, for one can say what goes wrong in the evil demon cases without thereby explaining what goes wrong in general in a Gettier case.

[4] See Ernest Sosa (1993) and (forthcoming). Indexical reliabilism is considered and rejected by Alvin Goldman (1991) and (1999) (see esp. pp. 10–11). "Indexical reliabilism" is my name for the position.

[5] See Robert Stalnaker (1978).

[6] "3:00 is the actual world" also contains an indexical, as we will see below, and so the same problem will arise regarding *its* truth-conditions; but I am trying to say what it is that, intuitively, Dave understands me as saying. It is difficult, if not impossible, to express that proposition without using indexicals, and that accounts for the ease with which we "diagonalize" in interpreting some utterances of sentences containing indexicals.

[7] Stalnaker (1978), p. 82. See also David Lewis (1973), pp. 61–64.

[8] I am not claiming that this is the application that Stalnaker had in mind in presenting the apparatus of propositional concepts, though. He explains what he had in mind in the "Introduction" to Stalnaker (1999).

[9] See David Kaplan (1989).

[10] An analogous problem would arise even for that kind of content externalism, however, under the guise of the problem of newly victimized subjects. Only an implausibly strong variety of content externalism would avoid the new evil demon problem in *any* of its versions.

[11] Modulo the generality problem, which we won't discuss here. See Earl Conee and Richard Feldman (1998).

[12] See David Lewis (1970), pp. 18–20. See also *Postcript* B to that paper, p. 22.

[13] Compare Gareth Evans (1979), pp. 201–2.

[14] David Lewis (1970), p. 19.

[15] *Reliabilism 1* is Sosa's V-ADROIT, and *Reliabilism 2* is Sosa's V-APT. Cf. Ernest Sosa (forthcoming). I am not sure what Sosa would say about the independece or connection between the two reliabilist notions, though what I say in the text is certainly in the spirit of his position.

[16] We can put some of the same points using Chalmers' terminology (David Chalmers, 1996). If I say "The demoners' beliefs are produced by reliable processes" what I say is true if we assign "reliable process" its *secondary intension*, but it is false if we assign to it its *primary intension*. Roughly put, the primary intension of a concept is a non-rigid function from possible worlds to extensions, whereas the secondary intension of a concept is the result of rigidifying the primary intension. Thus, the primary intension of the concept *water* is a function whose value at a world *w* is whatever fills the ocean and lakes and it is clear and drinkable *in w*. The secondary intension of *water* is a function whose

value at a world *w* is whatever fills the ocean and lakes and it is clear and drinkable *in the actual world*; that is, the secondary intension of *water* is a function whose value at *any* world is H_2O. Analogously, the primary intension of the concept *reliable process* is a function whose value at a world *w* is the set of processes that would produce mostly true beliefs *in w*. The secondary intension of *reliable process* is a function whose value at a world *w* is the set of processes that would produce mostly true beliefs *in the actual world*; that is, the secondary intension of *reliable process* is a function whose value at *any* world is {perception, memory, introspection, . . . }.

[17] As far as I know, they are not really located in the actual world, but rather in a possible world that is just like ours except that, in the year 3002, a demon victimizes everyone. But the problem still remains, for what will I say of the people in that other possible world in 3002? Are their beliefs diagonally justified or not?

[18] These "mixed" cases create a different kind of problem. Suppose that there are demoners in your neighborhood, and that you know about them. Does that threaten any of *your* knowledge? I don't think it does, although I will not defend the claim here. For an argument that in certain special cases it does, see Adam Elga (MS).

[19] The fallacy that I diagnose in TFA is similar in some respects to the fallacy that David Lewis (1970) finds in Anselm's ontological argument.

[20] This is, I believe, a fair, if rough, assessment of Richard Foley's starting point in Foley (1993).

[21] More precisely, what is necessary for warrant, if the reflections in section 3.3 are right, is reliability in their world and at their time and place. In what follows, this qualification will be left implicit.

[22] I am assuming that, although diagonal reliability is not sufficient for warrant, it is nevertheless a central necessary condition. In that sense, an indexical notion of reliability can "characterize" both justification and warrant.

REFERENCES

Chalmers, D. (1996): *The Conscious Mind*, Oxford: Oxford University Press.

Cohen, S. (1984): 'Justification and Truth', *Philosophical Studies* 46, 279–295.

Conee, E. and Feldman, R. (1998): 'The Generality Problem for Reliabilism', *Philosophical Studies* 89(1), 1–29.

Elga, A. (MS): 'Defeating Dr. Evil with Self-Locating Belief', unpublished.

Evans, G. (1979): 'Reference and Contingency', *The Monist* 62(2). Reprinted in Evans (1985), 178–213.

Evans, G. (1985): *Collected Papers*, New York: Oxford University Press.

Foley, R. (1993): *Working Without a Net*, New York: Oxford University Press.

Goldman, A. (1991): 'Epistemic Folkways and Scientific Epistemology', in his *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT/Bradford.

Goldman, A. (1999): 'A priori warrant and naturalistic epistemology', *Philosophical Perspectives, vol. 13: Epistemology*, pp. 1–28.

Kaplan, D. (1989): 'Demonstratives', in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, New York: Oxford University Press, pp. 481–563.

Lewis, D. (1970): 'Anselm and Actuality', *Nous* 4. Reprinted with Postscripts in Lewis (1983), 10–20.

Lewis, D. (1973): *Counterfactuals*, Cambridge, MA: Harvard University Press.

Lewis, D. (1983): *Philosophical Papers Volume I*, New York: Oxford University Press.

Sosa, E. (1991): *Knowledge in Perspective*, New York: Cambridge University Press.

Sosa, E. (1993): 'Proper Functionalism and Virtue Epistemology', *Nous* 27(1), 51–65.

Sosa, E. (forthcoming): 'Goldman's Reliabilism and Virtue Epistemology', *Philosophical Topics*.

Stalnaker, R. (1978): 'Assertion', in P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, New York: Academic Press. Reprinted in Stalnaker (1999), pp. 78–95.

Stalnaker, R. (1999): *Context and Content*, New York: Oxford University Press.

Brown University
Providence, Rhode Island
E-mail: Juan_Comesana@Brown.edu