Episteme, 9, 4 (2012) 345–360 © Cambridge University Press
doi:10.1017/epi.2012.26

THREE FORMS OF INTERNALISM AND THE NEW EVIL DEMON PROBLEM

ANDREW MOON
AndrewMoon@ksu.edu

ABSTRACT

The new evil demon problem is often considered to be a serious obstacle for externalist theories of epistemic justification. In this paper, I aim to show that the new evil demon problem ('NEDP') *also* afflicts the two most prominent forms of internalism: moderate internalism and historical internalism. Since virtually all internalists accept at least one of these two forms, it follows that virtually all internalists face the NEDP. My secondary thesis is that many epistemologists – including both internalists and externalists – face a dilemma. The only form of internalism that is immune to the NEDP, strong internalism, is a very radical and revisionary view – a large number of epistemologists would have to significantly revise their views about justification in order to accept it. Hence, *either* epistemologists must accept a theory that is susceptible to the NEDP *or* accept a very radical and revisionary view.

1. INTRODUCTION

Let *reliabilism* be the view that a belief is justified if and only if it is formed by a reliable process. Keith Lehrer and Stewart Cohen (1983: 192–3) have produced the following well-known and powerful counterexample to reliabilism:[1]

> 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 … 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.

Ralph Wedgwood (2000: 349) uses the demon hypothesis to support *epistemic internalism*:

[1] Other early presentations are found in Cohen (1984), Ginet (1985: 178) and Foley (1985: 190–1). Responses are found in Bach (1985), Goldman (1986, 1988), Sosa (1991), Engel (1992), LittleJohn (2009) and Lyons (forthcoming).

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

> Consider two possible worlds, w_1 and w_2. In both worlds, you have exactly the same experiences, apparent memories, and intuitions, and in both worlds you go through exactly the same processes of reasoning, and form exactly the same beliefs. In this case, it seems, exactly the same beliefs are rational in both worlds, and exactly the same beliefs are irrational in both worlds. Now suppose that in w_1 you are bedeviled by an evil demon who ensures that many of your experiences are misleading, with the result that many of the beliefs that you hold in w_1 are false. In w_2, on the other hand, almost all your experiences are veridical, with the result that almost all the beliefs that you hold in w_2 are true. Intuitively, this makes no difference at all. Exactly the same beliefs are rational and irrational in both worlds … This intuition seems to support an 'internalist' conception of rational belief. According to this conception, the rationality of a belief supervenes purely on 'internal facts' about the thinker's mental states – in this example, on facts that hold in *both* these two possible worlds w_1 and w_2, not on facts about the external world that vary between w_1 and w_2.

The *old* evil demon problem used the demon hypothesis in an argument for skepticism. The *new* evil demon problem uses it both as a counterexample to reliabilism and also as an argument for internalism and against externalism. It has been very influential. Ernest Sosa (1991: 131) identifies it as one of the 'three main problems' for generic reliabilism, Laurence BonJour (2002: 246) identifies it as one of the 'three main objections' to reliabilism, and Alvin Goldman (1992: 435) lists it as one of the two 'most influential counterexamples to reliabilism'.

In this paper, I aim to show that the new evil demon problem (from now on, NEDP) *also* afflicts the two most prominent forms of internalism: moderate internalism and historical internalism. Since virtually all internalists accept at least one of these two forms, it follows that virtually all internalists face the NEDP.[2] My secondary thesis is that many epistemologists – including both internalists and externalists – face a dilemma. The only form of internalism that is immune to the NEDP, strong internalism, is a very radical and revisionary view. This is not to say that strong internalism is *implausible* but only that a large number of epistemologists would have to significantly revise their views about justification if they accepted it.[3] Hence, *either* epistemologists face the NEDP *or* they must accept a very radical and revisionary view.

---

[2] A few epistemologists have noticed that the NEDP applies to certain versions of internalism. Lyons (forthcoming: section 5) has noticed that the NEDP afflicts certain weak versions of internalism. Huemer (1999) has noticed that it afflicts a view similar to a version of internalism. (For more on Huemer, see §4.2.) The *problem of forgotten evidence*, discussed by Alvin Goldman and others, might be seen as a precursor to my NEDP for internalism; it is supposed to show that certain memorial beliefs are not justified by any nonoccurrent beliefs (since they are forgotten). My arguments are more radical; they conclude that *no* nonoccurrent belief is directly relevant to the justification of *any* belief. (For more on the problem of forgotten evidence, see §4.1.) Sidelle (2001) has argued that the NEDP applies to any internalist view that allows for fallible justificatory conditions. A virtue of my arguments over Sidelle's is that they do not appeal to the difficult notion of *infallibility*, which becomes a sticking point in the discussion of Sidelle's argument in Conee (2002) and Sidelle (2002). Lastly, Sanford Goldberg (forthcoming) argues that the NEDP can show that the property which turns unGettiered, true belief into knowledge is not internalist (i.e. it does not supervene on internal properties of the believer). Thanks to the reviewers and editors of *Episteme* for pressing me to make note of these other arguments.

[3] Hence, one could take the arguments of this paper to all be one long argument in favor of strong internalism.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

I will end §1 by making explicit both whom I am trying to persuade and also what I am taking myself to accomplish. My main audience is internalists and any epistemologists, whether internalist or externalist, who share internalist intuitions – those who feel the temptation to say that the demon victim has justified beliefs just as we do. Externalists who do not feel this temptation will likely not feel moved by my arguments and are not my primary audience. Still, these externalists should be interested to know that an argument, which a significant number of their epistemologist colleagues regard to be a significant threat to externalism, also applies to the two most prominent forms of internalism. Second, I do not take myself to be presenting knockdown arguments against these forms of internalism. Just as some externalists might think that other theoretical considerations in favor of externalism outweigh the NEDP against it, some internalists might think the same. I am only attempting to show that the problem does apply to these forms of internalism, whether it is knockdown or not.

2. INTERNALISM, EXTERNALISM AND THE NEDP

In this section, I will define some terms. *Internalism* is the view that justificational properties supervene on internal properties of the believer.[4] In other words, only internal properties are directly relevant to whether a belief is justified. I say *directly* relevant, because any internalist will grant that external properties can be *indirectly* relevant to a belief's justificational status by affecting the internal properties. But only internal properties directly make a justificational difference.[5] Externalism, as I will understand it, is the denial of internalism. It entails the possibility that there are two individuals with identical internal properties but different justificational properties. The difference in justification would be a direct result of a difference in some external property, such as whether the belief was formed reliably.

What is it for a property to be *internal*? Perhaps *T* is internal if and only if *T* is *mental*. But Timothy Williamson (2000) has argued that knowledge is a mental state, and most internalists would not regard knowledge as internal. It would be preferable to have a definition of 'internal' that did not depend on the failure of Williamson's argument. Here are two more plausible candidates. Perhaps *T* is internal if and only if *T* is a *nonfactive* mental state. Or perhaps *T* is internal to *S* if and only if *T* is introspectively accessible to *S*. For this paper, I will go with the former definition, simply because it is the more inclusive view.[6] Fortunately, an exact analysis of internality is unnecessary for the purposes of this paper, and an intuitive understanding should be enough.

---

[4] Bach (1985), Engel (1992) and LittleJohn (2009) distinguish between personal justification (i.e. a *person*'s being justified *in* a belief) and doxastic justification (i.e. a *belief*'s being justified) and say that the demon victim has personal justification but not doxastic justification. Which sort of justification am I talking about? Since most internalists themselves are not likely to make this distinction (and might even find it problematic), I will not make the distinction in their favor and will simply talk about the justificational status of a belief. However, those who follow these philosophers can simply substitute personal justification ascriptions for any of my justification ascriptions. They will see that my dilemma can apply to those who are internalists only about personal justification.

[5] Here, I am following Gibbons's (2006: 20) way of characterizing internalism.

[6] I thank a reviewer of *Episteme* for making this clear to me. However, I actually prefer the latter definition over the former for the reasons given by Bergmann (2006: ch. 3).

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

What, exactly, is the NEDP? In its statement by Lehrer and Cohen, the NEDP took the form of a *counterexample* against reliabilism; this is how it is normally used. But we can also take it, following Wedgwood, as showing that no external property is *directly relevant* to justification. The NEDP would thereby show not only that *being reliably formed* is *unnecessary* for justification but also that it is *irrelevant* to it; reliable belief formation is a fifth wheel that does no justificatory work. Following Wedgwood, I understand a set of cases to exemplify the NEDP for a view *V* if they meet the following conditions: (1) *V* entails that some property *P* is directly relevant to justification; (2) in one case, some person *S* has a belief with a certain justificational status, and *S* (or *S*'s belief) exemplifies *P*; (3) in another case, some person *S** is exactly like *S*, except for any differences entailed by the fact that, due to the work of a demon, *P* is not exemplified by *S* (or *S*'s belief); (4) there is the intuition that the justificational status of the beliefs of *S* and *S** is the same; and (5) these cases can be generalized. *P* is normally the property *being reliably formed*, but it could in principle also be other external properties (e.g. *being sensitive* or *being safe*), as well as *internal* properties. This opens up at least the possibility that the NEDP could also apply to internalism.

Some might object to my liberal use of 'NEDP' and say that it should only be used as a counterexample to reliabilism or as an argument against externalist theories. I disagree; I think that cases that meet my above conditions are sufficiently similar to the original new evil demon cases to merit being called 'instances of the NEDP'. How shall we resolve this dispute? Fortunately, nothing important stands on this issue. My objectors can take my first thesis to say that a problem that is *similar to* (but not the same as) the NEDP afflicts the two most prominent forms of internalism. This still leaves a serious challenge for virtually all internalists, which is what I care about most.

3. FIRST FORM OF INTERNALISM: MODERATE INTERNALISM

3.1 Defining moderate internalism

Let *moderate internalism* be the view that internalism is true and some unaccessed, internal properties can be directly relevant to the justificational status of beliefs. Stated as a supervenience thesis, it is the view that justificational properties supervene on internal properties, and it is not the case that justificational properties supervene on accessed properties. On moderate internalism, both accessed internal states (such as my *being appeared to redly*) and unaccessed internal states (such as a dispositional belief) can make the difference as to whether a belief is justified.[7]

I understand a state *T* to be *accessed* by a person *S* if *T* is present to *S*'s consciousness.[8] My being appeared to redly and my belief that *I am typing a sentence* (as I currently type this sentence) are accessed internal states. For *T* to be accessed by *S*, *S* does not need to have beliefs about *T*. For example, my being appeared to redly can be an accessed internal state even if I have not formed any belief about my being appeared to redly. Unaccessed

[7] In this paper, I will refer to internal and external *states* as well as properties; any differences between them are nugatory.

[8] I do not also say 'only if' because I am open to Gibbons's (2006) suggestion that some external states are accessible. Those who disagree could add 'only if' and have a complete definition.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

internal states might include beliefs and memories that I am not currently thinking about. If asked why I believe Fred ate the last cookie, I can quickly bring to mind my memory of Fred eating the last cookie. This memory was an unaccessed internal state prior to my being asked about Fred and the cookie.

How do these terms relate to the common occurrent/dispositional distinction in epistemology? I take it that *S*'s belief is occurrent if and only if the belief is present to *S*'s consciousness; a belief is dispositional if and only if it is not occurrent.[9] Hence, a belief (or desire or memory) is occurrent if and only if it is accessed, and it is dispositional if and only if it is unaccessed. For example, my belief that $1 + 2 = 3$ is currently occurrent (and accessed); but I continue to believe this even when I am asleep. During those times, the belief is dispositional (and unaccessed). (I also take 'background belief' and 'stored belief' to both be synonymous with 'dispositional belief'; I will use these expressions interchangeably.)

Having explained my terminology, I think it is safe to say that virtually all internalists are moderate internalists. Otherwise, the set of potentially justifying (or defeating) beliefs would be severely restricted since only a relatively small number of our beliefs are occurrent at any given moment. So, most internalists will allow for nonoccurrent beliefs to be directly relevant to the justification of a belief. Hence, most internalists are moderate internalists.[10]

3.2 *The global NEDP for moderate internalism*

Consider Augustine, who lived for seventy-six years. During this time, he held a great many beliefs; some were justified and some were not. Let us only consider the beliefs when they were occurrent. Now imagine Augustine*, a creature who also lives for seventy-six years and who experiences all of the same *accessed* internal states that Augustine had throughout his lifetime. The only difference is that the demon is manipulating Augustine* in such a way that he never has *any* unaccessed internal states. Augustine* might *think* that he has dispositional beliefs at various times in his life, but, in fact, he has none. Whenever he thinks that he is recalling one of his memory beliefs, the demon is actually creating a newly formed occurrent belief. This scenario seems metaphysically possible. Furthermore, it seems that the justificational status of Augustine's occurrent beliefs at any age X of his life will be identical to the justificational status of Augustine*'s parallel beliefs when he is also age X, *even though* Augustine* has *no* unaccessed internal states.[11]

Consider that *you*, at this moment, might be fooled by a demon so that you are in a position similar to that of Augustine*. You might *think* that you have dispositional beliefs and memories, but you don't. You might try to prove that this isn't so by saying, 'Look, I

[9] Here I was helped by Senor's (1993: 461–2) definitions: 'A belief is occurrent at t iff it is conscious at t … S's belief that P is dispositional at t iff at t, S believes that P and P is not occurrent.'

[10] More support for the prevalence of moderate internalism will be made in §5.

[11] As I tell the story, it is implied that the demon is *intentionally* making Augustine*'s accessed states like Augustine's. But this is not an essential feature of the story. We can imagine Augustine**, who is also a victim of a demon so that his accessed internal states are like Augustine*'s, but this occurs by chance because the demon's actions are always on a whim. Still, it seems Augustine**'s beliefs have the same justificational status as those of Augustine and Augustine*. Note that internalists who reject this point will have to accept the view that the intentions of a demon a million miles away (or nowhere, if the demon is not spatially located) can be directly relevant to justification. Such a factor is about as external as you can get and internalists should not be happy with such a result.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

can bring to mind my dispositional belief that my childhood home was yellow.' Unfortunately, you could still be duped by a demon. Whenever you try to bring this belief to mind, the demon could quickly create a brand-new occurrent belief. Trying to prove that the demon is not fooling you about your dispositional beliefs by *recalling* them is as futile as trying to prove that the demon is not fooling you about external objects by *touching* them. It would be to no avail. Now I do not think that you are actually fooled by a demon, but there is a possible world in which you have a twin, You*, who *is* in such a situation. And it seems that the justificational status of your occurrent beliefs and those of You* is the same.

Consider yet another example from the movie the *Matrix*. Many think that Neo, before he was freed from the Matrix, had justified beliefs. Suppose that Neo had a counterpart, Neo*, who was phenomenologically identical to Neo. It would seem to Neo* that he was typing into a computer, it would seem to him that he was recalling memories, and so on; they would share all the same sensory experiences. However, suppose Neo* had no unaccessed, internal states because the Matrix determined it to be so. It seems that the beliefs held by both Neo and Neo* would have the same justificational status. These three cases give us evidence that unaccessed internal states are not directly relevant to the justificational status of our occurrent beliefs.

Someone might object that my cases do not show that unaccessed internal states are not directly relevant to the justificatory status of *dispositional* beliefs. Suppose Augustine has two dispositional beliefs B1 and B2. Perhaps B2 is directly relevant to B1's justificatory status. In response, someone who thinks this *and also* agrees with my argument that unaccessed, internal states are not directly relevant to the justificational status of occurrent beliefs must think that if B1 *were* to become occurrent, then B2 would suddenly become irrelevant to B1's justificatory status. But this seems implausible. It seems more reasonable to think that B2 is not directly relevant to the justificatory status of B1 in the first place. So, we have reason to think that unaccessed, internal states are not directly relevant to the justificatory status of *any* of our beliefs.

Consider the following objection. Some might think that these cases are impossible for at least one of two reasons: wide-scale deletion of *all* of a person's unaccessed internal states will either alter the person's accessed internal states or result in the loss of the person; hence, my cases are impossible. Yet, it does still seem to me to be metaphysically possible that such deletions could happen because it seems possible that there is someone who is exactly like me but has only my accessed states. So, I take the claim that my cases are metaphysically impossible to be insufficiently grounded.

My objectors might not budge and insist that they are impossible. Instead of pursuing this discussion, I will turn to another way of pressing the NEDP for moderate internalism. The current section employed what I call the *global* NEDP, cases in which *S* and *S** are internally identical except that *all* of *S**'s unaccessed internal states are deleted by a demon. The following section employs the *local* NEDP, which only makes use of cases in which a limited number of specific, unaccessed internal states are deleted; these cases are immune to the above objections.

3.3 The local NEDP for moderate internalism

Moderate internalists normally have in mind certain cases in which unaccessed, internal states matter. Consider the following:

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

(Case 1) Fred has been outside working all day. He looks out over the horizon and has a rich sensory experience as of a field, of trees, of a drifting cloud, and of the sun. Fred forms the belief that *that's a beautiful sunset*. He also has the dispositional belief that *it is evening*.
(Case 2) Sally wakes up and she has no idea how long she's been sleeping. She considers that she may have been sleeping anywhere from three to twelve hours. Sally looks out into the horizon and has a rich sensory experience as of a field, of trees, of a drifting cloud, and of the sun. Sally forms the belief that *that's a beautiful sunset*. Sally has no dispositional belief that *it is evening*.[12]

Neither Fred nor Sally has any background beliefs about which directions are east and west, and they cannot distinguish sunrises from sunsets by way of mere visual appearance. Furthermore, although Fred's belief that *it is evening* is dispositional, it plays a causal role in the formation of his belief that *that's a beautiful sunset*. It seems that Fred's belief about the sunset is justified and Sally's is not. If we asked Fred how he knows it is a sunset and not a sunrise, Fred would bring to mind some memory beliefs and say that sunrises do not occur in the evening, and it is evening. If we asked Sally the same question, she would say, 'Oh, I didn't consider that it might be a sunrise. It looks like I judged too quickly!'

Furthermore, as I have explained the cases, the accessed internal states of Fred and Sally are the same when they form their beliefs about the sunset. They both look out over the horizon, they are appeared to in a certain way, and they form the belief. Although Fred's belief that *it is evening* plays a causal role in the formation of his belief that *that's a beautiful sunset*, he does not bring it to mind. The only relevant differences between Fred and Sally's internal states are their dispositional beliefs and memories. Moderate internalists will likely point to these states as the best explanation for the justificational difference between them.[13]

However, other considerations might show that unaccessed, internal states are not actually doing the justificatory work. Consider:

(Case 3) Melissa has been outside working all day. She looks out into the horizon and has a rich sensory experience as of a field, of trees, of a drifting cloud, and of the sun. Melissa forms the belief that *that's a beautiful sunset*. She also has the dispositional belief that *it is evening*. An evil demon intervenes. First, he destroys Melissa's background belief that it is evening, and he himself causally sustains Melissa's occurrent belief that *that's a beautiful sunset*. After three seconds, the demon manipulates Melissa's mind so that (1) she once again has the background belief that *it is evening* and (2) the belief that *that's a beautiful sunset* is once again causally dependent on the background belief and not the demon. The demon is so skilled that throughout this process, Melissa's accessed internal experience remains uninterrupted as she enjoys the beautiful sunset.

Melissa's accessed states are exactly like Fred's. Furthermore, like Fred, Melissa's dispositional belief that *it is evening* plays a causal role in the formation and sustaining of her belief that *that's a beautiful sunset*. The only differences between them are the disappearance and reappearance of the background belief and the demon's causally sustaining her belief that *that's a beautiful sunset* when her background belief is absent. Plausibly,

[12] These examples are inspired by Senor (2005).

[13] Plantinga (1993: 99–101) thinks that a large number of our perceptual beliefs are based on background beliefs in this way. He writes: 'In order to be able to see that something is an orange, in order to be able to form the judgment *I see an orange* in the usual way, I must also know or take it for granted that *things that look like* ***that*** are oranges. A Maasai tribesman can't see that something is a 1986 Chevrolet, and someone just off the plane from Chicago can't see that a rhinoceros has recently passed by' (100); he gives a number of other examples.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

Melissa's belief that *that's a beautiful sunset* remains justified throughout the whole process.[14] Furthermore, Case 3 seems possible; a sufficiently talented and powerful demon could perform the task of deleting and creating a person's background belief while keeping her accessed mental states the same.[15] This possibility is easier to see than the possibility of Augustine* because it is easier to imagine the deletion of a limited number of unaccessed states rather than *all* of them.

Some may object that when Melissa's background belief disappears, it seems to them that her perceptual belief is not justified. I have two responses. First, note that if Melissa's belief is not justified because of the absence of the dispositional belief, then she epistemically ought to suspend judgment during those three seconds. This is counterintuitive.[16] Second, consider that *you* might be subject to a demon's tinkering in this way. The dispositional beliefs that you are inclined to think are relevant to the justification of your ordinary perceptual beliefs might be regularly deleted and recreated by a demon. It might be that in the last five minutes, you should have withheld some of your perceptual beliefs in increments of three seconds. Now, I do not think that you are actually in this situation, but your twin You** is. And it seems that the justificational status of your beliefs and You**'s beliefs are the same, regardless of those dispositional beliefs.

We can generalize from these examples. For any one of some subject *S*'s occurrent beliefs that we are inclined to think is justified (or unjustified) from t_1 to t_2 at least partly in virtue of some dispositional mental state *M*, we can imagine a counterpart *S** who is identical to *S* except for any differences entailed by the fact that *M* is deleted by a demon at t_1 and recreated at t_2. Like in the Melissa and You** cases, my intuition about any such case will likely be that the justificational status of *S* and *S**'s occurrent beliefs from t_1 to t_2 is the same. It follows that unaccessed, internal properties are not directly relevant to justificational properties, and so moderate internalism is false.

In §3, I do not consider myself to have *refuted* moderate internalism. As I said in §1, other theoretical considerations could still tip the balance of evidence in favor of moderate internalism. But I believe I have shown that the NEDP (or something similar to it) is a problem for moderate internalism; it is a problem worth taking seriously. And since virtually all internalists are moderate internalists, I believe I have shown that there is a problem for the views of virtually all internalists.

4. SECOND FORM OF INTERNALISM: HISTORICAL INTERNALISM

4.1 *Historical internalism*

Let *historical internalism* be the view that internalism is true and some nonpresent internal properties can be directly relevant to justificational properties. Stated as a supervenience thesis, it is the view that justificational properties supervene on internal properties, and

[14] We can stipulate that the demon has also deleted any other dispositional memories or beliefs that the accessibilist internalist might think is relevant, e.g. the belief that *the day has gone by* or that *it will be night soon*. I am focusing on the belief that *it is evening* for ease of discussion.

[15] Perhaps the deletion of dispositional beliefs is easier to imagine than their creation. Notice that I need only the disappearance of the belief to show the justificational irrelevance of Melissa's dispositional belief to her occurrent belief. However, since I think it is also plausible that a demon could create such beliefs, I will leave my description of the case as it is.

[16] Thanks to Clayton LittleJohn for this point.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

it is not the case that justificational properties supervene on present internal properties. One might object that historical internalism is not genuinely *internalist* because past internal states, though they are internal, are inaccessible *now*.[17] I would rather not engage in a dispute of how 'internalism' should be used. Instead, I will simply say that some internalists *do* think of historical internalism as a form of internalism, and I will argue that the NEDP applies to it, whether it is genuinely internalist or not.[18]

Why might someone accept historical internalism? Gilbert Harman (1986: ch. 4) points out that *forgotten evidence* might play a role in the current justification of our beliefs. Following Harman, Thomas Senor gives the following example:

> Harman notes that people often forget the original grounds of their beliefs even though, intuitively, those beliefs remain justified (for instance, I take myself to be justified in believing that my first-grade teacher's name was Mrs. McDonald, but I don't have any idea how I came by this belief, nor do I take myself to now have any other evidence for it). (Senor 1993: 454)

This might move some philosophers to think the following:

Past Justification Principle: Justification *S* had for *p* at past time t_1 can be directly relevant to *S*'s justification for believing that *p* at later time t_2.[19]

On the other hand, John Greco (2005, 2009) and Alvin Goldman (2009) both argue that past *defeaters* can be directly relevant to the justificational status of a present belief. Here is a recent example from Goldman (2009: 324).

> Ursula acquires by testimony an initially justified belief in a certain generalization, *G*. This occurs at t_1. She subsequently encounters a mountain of counterexamples to *G*, each of which is very obvious. The joint effect of these counterexamples, at t_5, is the total defeat of Ursula's original justification for *G* . . . hence she is unjustified in believing it. Nonetheless, Ursula continues to believe *G* . . . Much later, at t_{20}, Ursula continues to maintain the same high confidence in *G* as she has held right along . . . This is a clear case in which Ursula's belief in *G* at t_{20} is unjustified. This assessment holds even if we also stipulate that at t_{20} Ursula has long since forgotten the defeating evidence she encountered at t_5.

Goldman probably stipulates at the end that Ursula *forgets* the defeating evidence because he wants to show that the *past* encounters of the counterexamples at t_5 do justificational (or defeating) work, not the *present* memories or dispositional beliefs about the counterexamples. These cases support the following:

Past Defeat Principle: Defeaters *S* had for *p* at past time t_1 can be directly relevant to *S*'s justification for believing that *p* at later time t_2.

---

[17] E.g. see Greco (2005) and Goldman (2009: 327–8).

[18] The internalists I have in mind are Feldman (2005) and Huemer (1999). Huemer is not explicit, but in the 1999 article, he endorses a version of historical internalism, and it is well known that Huemer self-identifies as an internalist. See also the helpful discussion about the definition of 'internalism' between Feldman (2005) and Greco (2005). I am inclined to agree with Feldman that there is no clear answer to this question since 'internalism' is partly a stipulatively defined term.

[19] Goldman (2009, 2011) appears to endorse this as well.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

Internalists who accept these principles should believe historical internalism.

A historical internalist can also explain the justificational difference between Fred and Sally in Cases 1 and 2 without appealing to unaccessed, internal states. Fred will likely have formed the occurrent belief that it is evening sometime *in the recent past* for it to have become a dispositional belief; Sally never formed this belief.[20] Even if moderate internalism is false, Fred's past accessed, internal state might make the justificational difference between him and Sally. And the Past Justification Principle explains why this is so. So, by embracing historical internalism, one can reject moderate internalism, retain the intuitions about Cases 1 and 2, and remain internalist.

4.2 *The global NEDP for historical internalism*

Unlike the hitherto unknown global NEDP for moderate internalism, the global NEDP for historical internalism has already been used by Michael Huemer.[21] Inspired by Bertrand Russell, he writes the following:

> But now recall Russell's five-minute hypothesis. Suppose God created someone five minutes ago in exactly the state that I was in five minutes ago, surrounded by exactly the same kinds of things. Call this person Mike2. Mike2 was created complete with false memories of his past life, identical to my memories of my past life. He thinks his name is 'Mike' and is presently writing a paper about the problem of memory knowledge. His situation would be (to him) indistinguishable from my actual situation … What sorts of things would it be rational for Mike2 to believe? Pretty clearly, just the same things that it is rational for me, now, to believe (modulo appropriate changes in indexical references). Most of Mike2's beliefs about his own past are false, but he has no way of knowing that, and no more reason for suspecting it than I have for suspecting that my beliefs about my past are false. So if I am justified in believing that I ate a bagel this morning, Mike2 is justified (though mistaken) in believing that he ate a bagel this morning. (Huemer 1999: 350)

Here is another example. Consider that it is *possible* that you have a twin, You***, who is a molecule-by-molecule replica of you. You*** has all of the same memories and beliefs that you do, but You*** was created five minutes ago. (Note that you have no way of confirming that you are not You***!) The justificational status of your beliefs and You***'s beliefs seem to be the same.

We can also consider an alternative version of Case 1 (call it Case 4) in which a demon created the world a few minutes before Fred's internally identical counterpart, Fred*, looks out into the horizon and forms the belief that *that's a beautiful sunset*. Fred* shares Fred's stored memories and beliefs, but Fred*'s are mostly false; he also has the background belief that *it is evening*. Given the presence of this background belief, it seems that Fred*'s belief

[20] What if he didn't form that occurrent belief in the recent past? Robert Audi (1994: 420–1) thinks that it is not unusual for humans to form a dispositional belief without that belief ever being occurrent. Still, since Fred was working outside all day, he will have formed many *other* occurrent beliefs in the past that can play the same role as the belief that *it is evening*.

[21] Huemer does not aim the NEDP specifically at historical internalism but at *preservationism*, the view that the justification for a memorial belief is the justification it had when it first formed. Preservationism is similar to historical internalism in that both say that past states can be directly relevant to a belief's justification.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

that's a beautiful sunset is justified. We can similarly formulate an alternative version of Case 2 (call it Case 5) in which a demon created the world a few minutes before Sally* formed her belief. Sally* shares Sally's stored memories and beliefs, but Sally*'s are mostly false; she also does not have the background belief that *it is evening*. Given the absence of this background belief, it seems that Sally*'s belief *that's a beautiful sunset* is not justified. It seems that the past does not make a direct justificational difference to Fred and Sally's beliefs. One might point out that what motivates these intuitions is that, when we imagine Cases 4 and 5, we imagine Fred* with unaccessed internal states (the background belief that it is evening) that Sally* does not have. Doesn't this bring us back to moderate internalism? I will discuss this question in §5. For now, I will say that it is clear that these sorts of cases can be multiplied. In reflecting on these three examples, it seems plausible that past internal properties are not directly relevant to justificational properties.

What, then, are we to say about the Past Justification Principle and the Past Defeat Principle? Or what should we say about the cases of forgotten evidence that motivated them? Recall that in wielding the NEDP against historical internalism, I am not trying to give a knockdown argument against it. I mainly just want to show that it is a problem worth taking seriously, and I acknowledge that there might be other theoretical considerations that make historical internalism an overall plausible view to hold.[22]

Some might object that these cases are impossible. For example, they may say that Mike2 has the same beliefs and intentional states as the real Michael Huemer only if Mike2 has a certain sort of history, one precluded by Mike2's being created five minutes ago. For example, perhaps a being has intentional states only if it has had the right sort of evolutionary history.[23] Or perhaps a being can have a dispositional belief only if it was occurrent at some point in the past.[24] Despite these objections, it does still seem to me to be metaphysically possible that a molecule-by-molecule replica of me (created five minutes ago) could share the same beliefs and memories I have, even dispositional ones.

My objectors might not budge and insist that they are impossible. Instead of pursuing this discussion, I will turn to another way of pressing the NEDP for historical internalism. The current section employed what I call the *global* NEDP, cases in which *S* and *S** are internally identical except that almost *all* of *S**'s past internal states are absent due to a demon. The following section employs the *local* NEDP, which only makes use of cases in which a limited number of specific, past internal states are absent due to a demon; these cases are immune to the above objections.

4.3 The local NEDP for historical internalism

Historical internalists normally have in mind certain cases in which past internal states are supposed to matter, cases like Goldman's and Senor's cases. Since Goldman's case is the

[22] With that said, in §4.3, I show how the support for these two principles might be undermined.

[23] Proponents of teleological theories of mental content hold that the presence of the biological functions necessary for intentionality are conferred by natural selection. However, it seems to me that a demon could also confer functions by way of intentional design. In any case, I only raise the above point as a possibility.

[24] Huemer (1999: 356, n. 15) holds this view. Its truth is crucial for Huemer's own theory of memory, but a discussion of it is beyond the scope of this paper.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

most detailed and described, I will focus on it.[25] In that case, Ursula justifiedly believed G at t_1 but encountered counterexamples to G at t_5. She continued to believe G until t_{20}, by which she had 'long since forgotten the defeating evidence she encountered at t_5'. Now, the historical internalist might suggest a case with Ursula* who justifiedly believed G at t_1, never encountered the counterexamples at t_5, and has the same internal states as Ursula at t_{20}. To make this follow the structure of the local NEDP for historical internalism, we can even say that a demon was the reason why Ursula* never encountered the counterexamples at t_5. If the local NEDP is successful, then we should expect the justificatory states of Ursula and Ursula* to be the same, thus showing that the difference in past internal states didn't matter. However, we instead find that it seems that Ursula*'s belief is justified at this later time, and Ursula's belief is not. Furthermore, the only relevant differences between them are their past internal states. Historical internalists will likely point to these states as the best explanation for the justificational difference between them. They may also point out that the local NEDP for historical internalism fails.

I will show both that Goldman's case, once clarified, fails to provide support for historical internalism and also that the local NEDP is vindicated. I will clarify three points about Goldman's case. First, we sometimes use the word 'forget' in such a way that a person might forget that *p* but still dispositionally believe or know that *p*. For example, I might say, 'I knew that we were meeting today at noon, but I forgot that when I left the house.' After being told his new acquaintance's name for the ninth time, Fred exclaims, 'No, I knew that your name's Norma! I just forgot for a second, and I needed a moment to remember it.' Goldman's use of 'forgotten' may not properly leave in the mind of the reader the impression that Ursula's relevant beliefs and memories are *completely* wiped out.

Second, notice just *how many* memories and dispositional beliefs must be erased. It cannot just be Ursula's belief that *in the example I just read, an instance of the antecedent of G is true and an instance of the consequent seems false*. Other beliefs such as *I read about a counterexample to G in that book* and *Smith and I talked about a counterexample to G the next day* and even beliefs such as that *I have at some time read about* some *putative counterexamples to G* might also be evidence against G. These many beliefs might be deeply stored in the mind, and they too must disappear.

Third, if Ursula is an ordinary human being, memories of the encounter with the counterexamples will linger for a very long time, and the many relevant memories of the events surrounding the encounter will last even longer. So, Goldman's use of the word 'forgotten', the sheer number of beliefs and memories that must be erased, and the natural assumption that Ursula is an ordinary human being may detrimentally affect the reliability of our intuitions about his example.

We must stipulate, therefore, that *all* of the potentially relevant internal states have been *completely erased* from Ursula's mind so that she is internally identical to Ursula* at t_{20}. (Since ordinary humans do not lose memories this easily, we can even stipulate that a demon wipes out all the relevant memories.) From her internal point of view at t_{20}, it is as if the event never happened. At that moment, she has her initial reasons for the belief, and all of the potentially defeating dispositional memories and beliefs are gone. But when the case is specified this way, it seems to me that, at t_{20}, she is epistemically

[25] What I say in response to this case can apply to the other cases of forgotten evidence or defeat as well. When you structure the cases so that you have two twins who were not internally identical in the past, but it is clear that they are internally identical *now*, it still seems that the two are equally justified.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

justified in believing G.[26] Hence, it seems that the justificational status of Ursula and Ursula*'s beliefs are the same at t_{20}.[27] They both justifiedly believe G, and Ursula's past encounter with the counterexamples seems justificationally irrelevant.

We can generalize. For any of some subject *S*'s present beliefs that we are inclined to think is justified (or unjustified) at *t* at least partly in virtue of some specific past internal state, *P*, we can imagine a counterpart *S** who is identical to *S* at *t* except for any differences entailed by the fact that *S** was never in *P* because of a demon. Given that *S* and *S** are completely internally identical at *t*, my intuition about any such case will likely be that the justificational status of *S* and *S**'s beliefs is the same. We can now see that the local NEDP for historical internalism is also worth taking seriously.

5. THIRD FORM OF INTERNALISM: STRONG INTERNALISM

Let us return to Cases 1 and 2. When I consider the global NEDP for historical internalism, I am inclined to think that unaccessed, internal properties make the justificational difference between Fred and Sally. When I consider the local NEDP for moderate internalism, I am inclined to think that past, accessed properties make the justificational difference between them. But both NEDPs combined seem to rule out any options for what could make the difference. If neither past accessed properties nor unaccessed, present properties are directly relevant, then how do we explain the justificational difference between Fred and Sally's beliefs?

Here is one answer: we don't. We accept that the justificational status of their beliefs is the same, and we take the lesson of the NEDP to be that only presently accessed, internal properties are directly relevant to justification. Richard Feldman (2004) has defended such a view. Briefly, Feldman first defended with his colleague Earl Conee (in Feldman and Conee 1985) the view that justification is solely a matter of the evidence one has. And in Feldman (2004: 219), he defended 'a restrictive account that limits the evidence a person has at a time to the things the person is thinking of or aware of at that time'. In other words, he defended *strong internalism*, the view that justificational properties supervene on presently accessed internal properties.[28] I will not explore Feldman's argument for strong internalism here; I only point out that the view has been independently defended, and it is an option that an internalist could take.

The strong internalist will probably deny that Sally's belief is unjustified. Fred and Sally both undergo the same perceptual experience; it also seems or appears to both of them that *that's a beautiful sunset*. Feldman thinks that experiences such as 'feelings of certainty' count as evidence for a belief; Fred and Sally would both share these experiences (2004: 239). Given that Sally has no accessed reason to doubt her belief, the strong

[26] My response in these paragraphs is inspired by moves made by Feldman (2000: 687–91). Comesana (2011) takes a similar line of reasoning.

[27] Remarks by Goldman (2009: 326) about a different version of the Ursula case imply that he would probably accept these intuitions as well.

[28] The name 'strong internalism' was coined by Goldman (1999: 278) for the view that 'Only facts concerning what conscious states an agent is in at time *t* are justifiers of the agent's beliefs at *t*.' I believe that this is equivalent to the view I am describing here as strong internalism.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

internalist will say that it seems that she *is* justified in believing that *that's a beautiful sunset*.[29] Furthermore, if someone were to ask her how she knows it is a sunset and not a sunrise, she might then form the occurrent belief that *I don't know whether what I am seeing is a sunset or a sunrise*. When she has this new accessed state, her belief that *that is a beautiful sunset* might then become unjustified, but not until then.[30] I do not find this explanation to be implausible. However, many will still consider this result to be a deal breaker. Some might think that if anything is true about justification, it is that Fred and Sally's beliefs are justificationally different. They will have a difficult time accepting strong internalism. I am not sure how to adjudicate this conflict.

I will end this section by defending my claim that strong internalism is radical and revisionary. First, consider that most epistemologists are foundationalists, coherentists or some mixture of the two. Coherentists will require that a belief be a member of a coherent set of beliefs in order to be justified; it is a very rare coherentist who restricts this set to presently occurrent beliefs. Foundationalists will require that a belief be based on other justified beliefs in order to be inferentially justified; it is a very rare foundationalist who restricts these justified beliefs to only presently occurrent ones. Someone who accepts a mixture of foundationalism and coherentism is also very unlikely to restrict the set of potentially justifying or defeating beliefs to only presently occurrent ones. We see that acceptance of strong internalism would require major revision in how many think about these traditional epistemological views, and the example with Fred and Sally only reveals one part of such a revision in our ways of thinking.[31]

Second, nobody in the literature explicitly *endorses* strong internalism, including Feldman, the one person who has defended it. In their latest work, Conee and Feldman (2011: 466–9) appeal to *dispositional states* in order to respond to one of Goldman's (2011) arguments. This commits Feldman to a rejection of strong internalism. Some may think that the phenomenal conservativist or dogmatist views of Michael Huemer (2001) and James Pryor (2000) are strong internalist views. Both views attribute prima facie justification to anybody with the appropriate seeming or appearance, which are accessed internal states. Doesn't that make them endorsers of strong internalism? To answer this, note that these views are only stating a *sufficient* condition for prima facie justification, i.e. justification absent defeaters. Neither Huemer nor Pryor have ever said that nonoccurrent beliefs cannot serve as defeaters. They might or they might not hold to this. Hence, their views are only *compatible* with strong internalism, but nothing they say commits them to it. So, as of yet, there is nobody on record who endorses strong internalism.

I am not attempting to argue here that strong internalism is false because it is very radical and revisionary. The main point of this section is that, although strong internalism is immune to the NEDP, many epistemologists would have to significantly revise their views in order to accept it. Perhaps, however, the arguments of this paper will move more

[29] Thanks to a reviewer of *Episteme* for help in defending the strong internalist view.

[30] See Feldman (2004: 239) for how he deals with a similar example. His first response mirrors the one I have presented here. His second response is basically to give up strong internalism, so I am not discussing it in the main text.

[31] For an argument against strong internalism, see Goldman (1999: 278). He says that the strong internalist must hold to the implausible view that our dispositional beliefs are justified by accessed internal states. See §A2 of Conee and Feldman (2001) for a response.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

philosophers toward accepting it. I myself am undecided on what I think is the truth regarding the matter.

6. CONCLUSION

I have shown that the sorts of properties that many internalists would like to appeal to – unaccessed, internal properties and past internal properties – are sufficiently *external* that a demon could delete (or could have deleted) them without the believer's conscious awareness. These internalists thereby face the same (or a very similar) NEDP that they have been leveling against externalists. As a result, we might be tempted to deny all of these views: externalism, moderate internalism and historical internalism. But this leaves us only with strong internalism, which will be hard for many to swallow. The result is a dilemma: either accept strong internalism or face the force of the NEDP.[32]

REFERENCES

Audi, Robert. 1994. 'Dispositional Beliefs and Dispositions to Believe.' *Noûs*, 28: 419–34.
Bach, Kent. 1985. 'A Rationale for Reliabilism.' *The Monist*, 68: 246–63.
Bergmann, Michael. 2006. *Justification without Awareness*. New York: Oxford University Press.
BonJour, Laurence. 2002. 'Internalism and Externalism.' In Paul Moser (ed.), *The Oxford Handbook of Epistemology*, pp. 234–63. Oxford: Oxford University Press.
Cohen, Stewart. 1984. 'Justification and Truth.' *Philosophical Studies*, 46: 279–95.
Comesana, Juan. 2011. 'Conservativism, Preservationism, Conservationism and Mentalism.' *Analysis*, 71: 489–92.
Conee, Earl, 2002. 'Innocuous Infallibility.' *Philosophy and Phenomenological Research*, 64: 406–8.
Conee, Earl, and Feldman, Richard. 2001. 'Internalism Defended.' In Hilary Kornblith (ed.), *Epistemology: Internalism and Externalism*, pp. 231–60. Oxford: Blackwell.
——. 2011. 'Replies.' In Trent Dougherty (ed.), *Evidentialism and its Discontents*, pp. 427–501. New York: Oxford University Press.
Engel, Mylan. 1992. 'Personal and Doxastic Justification in Epistemology.' *Philosophical Studies*, 67: 133–50.
Feldman, Richard. 2000. 'The Ethics of Belief.' *Philosophy and Phenomenological Research*, 60: 667–95.
——. 2004. 'Having Evidence.' In Earl Conee and Richard Feldman (eds), *Evidentialism: Essays in Epistemology*, pp. 219–41. Oxford: Oxford University Press.

[32] Thanks to Andrew Bailey, Michael Bergmann, Sanford Goldberg, John Greco, Clayton LittleJohn, Peter Markie, Kevin McCain, Matthew McGrath, George Pappas, Ted Poston, Philip Swenson and Chris Tucker for helpful comments on earlier drafts of this paper. Thanks to Earl Conee, Richard Feldman, Richard Fumerton, Jonathan Kvanvig, and Ernest Sosa for the helpful and encouraging discussion. This paper originally arose out of a heated discussion with internalists Kevin McCain and Philip Swenson; special thanks are due to them. Thanks also to the audiences at the 2009 Berkeley-Stanford-Davis Graduate conference, the 2010 Central APA, and the 2010 'Justification Revisited Conference' at Université de Genève, for their helpful questions and objections; thanks to my commentators at the former two of those conferences, respectively, Nate Smith and Matthias Steup. Thanks to the University of Missouri Chancellor's Dissertation Completion Fellowship, which allowed me to work on this paper. Thanks to numerous friends for help in picking the title of this paper. Lastly, thanks to Alvin Goldman and the reviewers of *Episteme* for helpful comments and advice.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press

——. 2005. 'Justification is Internal.' In Matthias Steup and Ernest Sosa (eds), *Contemporary Debates in Epistemology*, pp. 270–84. Malden, MA: Blackwell.
Feldman, Richard, and Earl Conee 1985. 'Evidentialism.' *Philosophical Studies*, 48: 15–24.
Foley, Richard. 1985. 'What's Wrong with Reliabilism?' *The Monist*, 68(2): 188–202.
Gibbons, John. 2006. 'Access Externalism.' *Mind*, 115: 19–39.
Ginet, Carl. 1985. 'Contra Reliabilism.' *The Monist*, 68(2): 175–85.
Goldberg, Sanford. Forthcoming. 'A Novel (and Surprising) Argument Against Justification Internalism.' *Analysis*.
Goldman, Alvin. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
——. 1988. 'Strong and Weak Justification.' *Philosophical Perspectives*, 2: 51–69.
——. 1992. 'Reliabilism.' In Jonathan Dancy and Ernest Sosa (eds), *A Companion to Epistemology*, pp. 433–6. Cambridge, MA: Blackwell.
——. 1999. 'Internalism Exposed.' *Journal of Philosophy*, 96: 271–93.
——. 2009. 'Internalism, Externalism, and the Architecture of Justification.' *Journal of Philosophy*, 106: 309–38.
——. 2011. 'Toward a Synthese of Reliabilism and Evidentialism?' In Trent Dougherty (ed.), *Evidentialism and its Discontents*, pp. 393–426. New York: Oxford University Press.
Greco, John. 2005. 'Justification is not Internal.' In Matthias Steup and Ernest Sosa (eds), *Contemporary Debates in Epistemology*, pp. 257–70. Malden, MA: Blackwell.
——. 2009. *Achieving Knowledge*. Cambridge: Cambridge University Press.
Harman, Gilbert. 1986. *Change in View*. Cambridge, MA: MIT Press.
Huemer, Michael. 1999. 'The Problem of Memory Knowledge.' *Pacific Philosophical Quarterly*, 80: 346–57.
——. 2001. *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.
Lehrer, Keith, and Cohen, Stewart. 1983. 'Justification, Truth and Coherence.' *Synthese*, 55: 191–207.
LittleJohn, Clayton. 2009. 'The Externalist's Demon.' *Canadian Journal of Philosophy*, 39: 399–434.
Lyons, Jack. Forthcoming. 'Should Reliabilists be Worried about Demon Worlds?' *Philosophy and Phenomenological Research*.
Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.
Pryor, James. 2000. 'The Skeptic and the Dogmatist.' *Nous*, 34: 517–49.
Senor, Thomas. 1993. 'Internalistic Foundationalism and the Justification of Memory Belief.' *Synthese*, 94: 453–76.
——. 2005. 'Epistemological Problems of Memory'. In *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/memory-episprob, 3 Jan. 2005 version.
Sidelle, Alan. 2001. 'An Argument that Internalism Requires Infallibility.' *Philosophy and Phenomenological Research*, 63: 163–79.
——. 2002. 'Innoculi Innocula.' *Philosophy and Phenomenological Research*, 64: 409–11.
Sosa, Ernest. 1991. 'Reliabilism and Intellectual Virtue.' In Ernest Sosa (ed.), *Knowledge in Perspective*, pp. 131–48. Cambridge: Cambridge University Press.
Wedgwood, Ralph. 2001. 'Internalism Explained.' *Philosophy and Phenomenological Research*, 65: 349–69.
Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

ANDREW MOON is a visiting assistant professor at Kansas State University. He specializes in epistemology and metaphysics. He also has interests in ethics and philosophy of religion.

https://doi.org/10.1017/epi.2012.26 Published online by Cambridge University Press