

SHOULD RELIABILISTS BE WORRIED ABOUT DEMON WORLDS?

Jack C. Lyons

University of Arkansas

March 5, 2011

The most straightforward version of reliabilism is one that links justification with reliability in situ, i.e., in the agent's world and/or environment. This "in situ reliabilism" contrasts with an actual world reliabilism, which interprets 'actual' rigidly, and connects justification to reliability in *this* world; it contrasts with an indexical reliabilism (Sosa 1993, Comesaña 2002), which combines an actual reliability dimension with an in situ dimension; and contrasts with normal worlds reliabilism (Goldman 1986), which appeals to reliability in worlds that satisfy our general beliefs about the world.

Besides bare simplicity, in situ reliabilism has the advantage of retaining a direct link between justification and truth. Actual world reliabilism, for example, and in situ reliabilism say the same thing about actual agents: their justified beliefs are objectively likely to be true. But a nonactual agent can be justified, on this theory, by using a process that is reliable here but not in her own world.¹ That belief would be justified but not likely to be true. Conversely, many of this agent's beliefs that result from reliable processes would therefore be likely to be true but nonetheless unjustified, merely because those processes are not reliable here.

One of the main reasons for investing in a more unwieldy reliabilism is the so-called New Evil Demon Problem, due to Lehrer and Cohen (1983). It is well-known by now, but worth quoting:

¹ I frame the discussion in terms of process reliabilism rather than indicator reliabilism. All the important claims should apply to the latter view as well.

Imagine that, unknown to us, our cognitive processes, those involved in perception, memory and inference, are rendered unreliable by the actions of a powerful demon or malevolent scientist. It would follow on reliabilist views that under such conditions the beliefs generated by those processes would not be justified. This result is unacceptable. The truth of the demon hypothesis also entails that our experiences and our reasonings are just what they would be if our cognitive processes were reliable, and, therefore, that we would be just as well justified in believing what we do if the demon hypothesis were true as if it were false. (p 192; see also Cohen 1984, p. 281)

In a demon world, all our cognitive processes would be unreliable, but so long as our experiences and reasonings were the same, our beliefs would be justified nonetheless; therefore, reliabilism---in situ reliabilism, at any rate---is false. Or so goes the objection.

The New Evil Demon Problem (henceforth, NDP) has been quite influential. Many epistemologists, internalist and externalist alike, think that NDP reveals a fatal flaw with in situ reliabilism, with internalists rejecting reliabilism altogether and reliabilists offering versions of reliabilism that appeal to something other than---or in addition to---reliability in the agent's world.² I think NDP poses no real threat to reliabilism, for I think the reliabilist is correct to hold that demonworlders (i.e., the victims of the deceiving demon, or malevolent scientist, or what have you) have very few justified beliefs.³

² Internalists who endorse NDP against reliabilism are too numerous to try to cite exhaustively. A few random examples are BonJour 2010, Feldman 2003, Haack 1993, Pollock and Cruz 1999; Huemer 2001 employs it not against reliabilism but in favor of his internalism. Reliabilists of various sorts have taken the argument quite seriously and offered modified versions of reliabilism accordingly. See Comesaña 2002, 2010, Goldman 1986, 1988, 1992, Henderson and Horgan 2001, Sosa 1993. I will take the NDP-wielding opponents of reliabilism to be internalists, in an undefined sense of the term. This is much less a serious attribution than the use of a convenient label.

³ My concern throughout will be with the in situ varieties of reliabilism. Consequently, and to reduce clutter, I'll drop the 'in situ' qualification, except for an occasional reminder. Any claims I make about reliability, if not explicitly marked to the contrary, should be read as claims about reliability in situ.

The plan is this: after some preliminaries in section 1, I will try to show that our demon world counterparts really are unjustified in many of their beliefs. To soften this claim, I argue in section 2 that the reliabilist can nevertheless attribute a positive epistemic status to many of the beliefs of the demonworlder; some but not all of these beliefs will actually be justified. This argument relies on the well-known distinction between belief-dependent and belief-independent processes (Goldman 1979); the former yield positive status even in demon worlds. Reliabilism withholds positive epistemic status only from beliefs that result from belief-independent processes. In section 3 I try to show that reliabilism is right to claim that such beliefs are unjustified. Because of the results of section 2, NDP is a threat to reliabilism only in the context of a view I call “Seemings Internalism,” (SI), which holds that having an experience as of p *prima facie* justifies one in believing that p (e.g., Pollock 1986, Pryor 2000). SI is false, however, and demon world considerations actually tell *against* it, not in favor of it. In section 4 I examine and reject two attempts to rescue SI from the arguments of section 3. In section 5 I consider some internalist theories that reject SI; I conclude that these theories do not support the NDP objection to reliabilism.

1. Preliminaries

It is well known that certain kinds of semantic externalism cause trouble for the very formulation of NDP. There’s a popular semantic theory that implies that if demonworlders are not in causal contact with chairs, then they can’t have beliefs about chairs and thus can’t have the same beliefs as I do (Putnam 1981). And there’s a respectable (though less popular) phenomenological theory that holds that if they’re not in causal contact with red things, then they

can't have experiences as of red and thus can't have the same experiential states as I do (Dretske 1996, Tye forthcoming).⁴ I propose to simply set these difficulties aside, not because I think the issues in the philosophy of mind don't deserve to be taken seriously, but because the demon world scenario provides a convenient framework for raising certain fundamental epistemological questions, questions whose importance remains whether the demonic deception is possible or not.

For example, demon scenarios render vivid and plausible the epistemic principle that any two agents who are experientially and doxastically identical are justificationaly identical as well. Lehrer and Cohen, in the above-cited passage clearly imply that if two agents have the same experiences and reasoning, then they are justified in the same beliefs. This is, or is very similar to, the view that Conee and Feldman (2004) call "mentalism," and it's important that we determine whether it is true. Semantic deflections of NDP fail to address this crucial epistemological worry. For these reasons, I will allow that demonworlders can have the same experiences, beliefs, etc. as the rest of us.

Second, I want to allow for a rather lax reading of 'in situ'; the process needs to be reliable *where it is*, but 'where it is' could be interpreted in more or less spatiotemporally restrictive ways. For example, a brain that is envatted for an hour and then replaced might count as having reliable cognitive processes and justified beliefs, even during the envatted period. This seems intuitively plausible and may even preserve the truth connection to an acceptable degree. There are important details to be worked out---we will have to decide whether reliability in the

⁴ Alternatively, one might endorse a neoBerkeleyan semantics for the secondary intension of 'chair' and the like, so that the demonworlders' beliefs are largely true. If we individuate beliefs by their primary intensions (which would be the same for us and them), we could end up with the demonworlders having the same beliefs as us, without those beliefs being false. Again, I will ignore this not because I think it deserves to be ignored but because I want to focus on the epistemological issues made salient by the ordinary take on NDP.

relevant sense is more a matter of track record, or future propensity, etc.---but they shouldn't really affect the present discussion, and I don't want to take a stand on them here. Whatever exactly reliability in situ turns out to be, it will be quite different from reliability in, say, normal worlds or our actual world.

Now, NDP comes in many varieties, and though I don't want to attempt a full taxonomy, two significant dimensions of variation should be considered. First, we can distinguish between what I will call evidential and dispositional versions of NDP. The Lehrer and Cohen version cited above is an evidential version: the problem for reliabilism is supposed to follow from the fact that the demonworlder has the same beliefs and experiences---in short, the same evidence---as someone whose beliefs are justified. The dispositional versions focus on other factors.

Henderson and Horgan's (2001) contrast between Faith and Constance appeals not to differences in evidence, but differences in how agents use the same evidence. Faith makes hasty generalizations, while Constance is very careful and is watchful for sampling biases and the like. It is not obvious whether such dispositions can be reduced to differences in evidence, and in case they cannot be, I'll treat dispositional versions of NDP as distinct from evidential versions. My focus will be on the evidential versions. I think my treatment of conditional justification below offers a response to the dispositional versions of the problem, but I won't try to argue that here. I claim only to be able to answer the evidential version.

The other important dimension of variation concerns the demands of the internalist: one variant of NDP complains that reliabilism can discern no difference at all between two subjects who are in fact epistemically different (e.g., two demonworlders, subjected to an intraworld comparison); another protests that reliabilism does discern a difference between two subjects

who are in fact epistemically identical (e.g., one of us and a demonworlder, subjected to an interworld comparison). (Again, contrast Henderson and Horgan 2001 with Lehrer and Cohen 1983, respectively.) My goal is to argue that reliabilism can, in fact, recognize epistemic differences among demonworlders, even though they have comparable numbers of true beliefs; but I also insist that there is an epistemic difference between us and demonworlders, and the demand that they come out to be epistemically identical is an excessive one. Thus, I will claim that the challenge to discern a difference, in the intraworld comparison, is a challenge that can be met, but the challenge to preserve perfect indistinguishability in the interworld comparison is one to which reliabilists should not rise. This latter challenge can be softened, to require not perfect epistemic identity but some positive epistemic status on behalf of the demonworlders; this challenge can, to a large extent, be met.

I view reliabilism as first and foremost a theory about *prima facie* justification. It is hoped that defeat and *ultima facie* justification can be understood in reliabilist terms as well, though that won't concern us here. In any case, NDP is about the reliabilist requirements for *prima facie* justification, not defeat. Subsequent discussion of justification should be understood to be concerned with *prima facie* justification. I will generally omit the qualifier for brevity.

Finally, I do not intend any of this as an argument *for* reliabilism. As we will see, there are residual problems for reliabilism, although these problems are different from those thought to be raised by NDP. My primary goal is to show that reliabilists can meet NDP without giving up the essential features of the view. A secondary goal is to show that certain popular internalist

theses---Seemings Internalism, mentalism, Evidence Essentialism---all face demon world difficulties of their own. Though it will fall far short of a general refutation of internalism, the discussion here should point out serious difficulties for certain brands of internalism.

2. Belief-Dependence, Conditional Reliability, and Conditional Justification

First, for the good news. I want to begin by arguing that in situ reliabilism can, in fact, attribute a positive epistemic status to (some) demonworlders' beliefs.

It is generally recognized, but worth making explicit, that even in a demon world, agents can have many true beliefs and thus some reliable cognitive processes. If rational intuition is reliable here, it's reliable there. And it presumably is reliable here, or at least certain subtypes or subprocesses of rational intuition are reliable here. Our epistemic counterparts in demon worlds will have rational intuition processes that are exactly as reliable as ours. And on the assumption that demonworlders really do have all the same experiential and doxastic states that we do (which I am taking for granted here), their introspective processes are as reliable as our own. So it is just not true that demonworlders don't have any true beliefs. And of course, anything deduced from these true beliefs will also be true, and so on.

But this brings me to a more important point, one which may also be generally recognized, though as far as I know, no one ever discusses it. Reliabilism can, and in some incarnations does, distinguish between belief-dependent processes (which count beliefs among their inputs) and belief-independent processes (which do not) (Goldman 1979). Perception, for instance, is generally thought to be a paradigm case of a belief-independent process, while deduction is a paradigm case of a belief-dependent process. Let us extend the terminology to

cover beliefs as well, so a belief-independent belief will be one that results from a belief-independent process, and so on. On Goldman's original (1979) reliabilist theory, a belief-independent belief is *prima facie* justified iff it's the result of a categorically reliable process (i.e., one with a high propensity to truth), while a belief-dependent belief is *prima facie* justified iff it's the result of a conditionally reliable process (i.e., one with a high propensity to truth, *given true beliefs as inputs*), all the doxastic inputs to which are themselves be justified. Deduction, since it is conditionally reliable even in demon worlds, will produce justified beliefs there when given justified inputs.

One quibble about NDP and then two serious points: the quibble is that at least one influential version of reliabilism does not, contrary to standard presentations (including Lehrer and Cohen's), require that memory and inference be reliable; because these are belief-dependent processes, it requires instead that they be *conditionally* reliable. But demonic interference only threatens categorical reliability. Thus, the standard presentations of NDP fail to address a standard version (indeed the *locus classicus*!) of reliabilism.

The first serious point to fall out of this quibble is that NDP should be thought of primarily as a problem concerning belief-independent beliefs. Memory and at least some types of inference *are* conditionally reliable, even in demon worlds, as much as they are here. If my demon world counterpart is psychologically identical to me, then his memory processes, deductive inference processes, etc. will be conditionally reliable if mine are. Although reliabilism charges demonworlders with vast numbers of unreliably produced beliefs, the bulk of the unreliability can be traced to the unreliability of the belief-independent processes, particularly perception. What makes the demonworlders' memory processes unreliable is just the fact that so

many of the doxastic inputs to these processes are false. The memory processes, however, are still doing their job---they are conditionally reliable. The problem, if there is a problem, is with the reliabilist epistemology of perception, not with the reliabilist epistemology of memory.

This requires a bit of elaboration. Memory is held by Goldman (1979) to be, as we might say, both epistemologically and psychologically belief-dependent. That is, not only is memory a process that takes beliefs as inputs (thus psychologically belief-dependent), but it is a process that, like inference, yields justified beliefs only if the beliefs given to it as inputs are themselves justified (thus epistemologically belief-dependent). The theory implies that all psychologically belief-dependent processes are epistemologically belief-dependent. One might deny this.⁵ One might instead think that reliabilism should construe memory as a generative, rather than merely preservative source of justification (see the debate between Lackey 2005, 2007 and Senor 2007). Either way, the (categorical) unreliability of memory in demon worlds derives from the unreliability of the “feeder” processes, not from memory itself. The internalist thinks that memory should not be epistemically downgraded simply because it is unreliable in demon worlds. The reliabilist agrees: the unreliability of memory is the fault of something other than memory; a good memory is one that is conditionally reliable, not necessarily one that is categorically reliable. Consequently, a reliabilist *could* drop the justified input requirement, demanding only conditional reliability, and maintain that memory beliefs are justified even in

⁵ There are actually three distinct properties that are held to be coextensive: psychological belief-dependence, epistemological belief-dependence, and the property of being the kind of process for which the justification conferring type of reliability is conditional, rather than categorical. I think Goldman (1979) is wrong to claim, as a general rule, that any process that is psychologically belief-dependent is epistemologically belief-dependent (Lyons 2009a, pp. 137--8): introspection, for instance, sometimes takes beliefs as inputs (as, e.g., when I introspect that I believe that p), but the epistemic status of my introspective belief does not depend at all on the epistemic status of the introspected belief. I need not be justified in believing that p, to be justified in believing that I believe that p. Similarly, even if certain perceptual processes are cognitively penetrable and therefore (psychologically) belief-dependent, the justificatory status of the resulting perceptual belief has little to do with the justification of the penetrating belief (for an argument, see Lyons forthcoming). In addition, it seems that we want to require of these processes (i.e., introspection and perception) categorical, rather than conditional, reliability.

demon worlds. Most reliabilists won't want to do this, but now it should be clear that the disagreement would concern the question of whether memory is epistemically belief-dependent, and this is orthogonal to the demon world concerns.

There is some categorical and conditionally reliability even in demon worlds. We can take this idea a bit further. Suppose that a roughly Cartesian epistemology adequately captured our perceptual psychology. That is, suppose that we formed perceptual beliefs by first introspecting our perceptual experiences, rationally intuiting the existence and perfection of the Deity, deducing from his perfection that God would not allow us to be massively mistaken about perceptual matters, and then concluding that things were (probably) the way our experiences represent them as being. All of the psychological processes just invoked---introspection, rational intuition, deduction---are categorically or conditionally reliable (as appropriate) in all worlds, including demon worlds. This is surely a large part of what Descartes liked about the corresponding epistemology. Perceptual belief would result from conditionally reliable belief-dependent processes that take justified beliefs as inputs. Perception would still be unreliable---most or all perceptual beliefs would be false---but the perceptual beliefs would be justified, according to a thoroughgoingly reliabilist epistemology. And this epistemology doesn't involve sophisticated new twists on the original reliabilist theory; *it is* the original reliabilist theory.

Of course, I don't think that the Cartesian story here is a remotely adequate account of the psychology of perception, but the important point is that if it were, a standard form of reliabilism would imply that our perceptual beliefs were justified, even in demon worlds.

The second serious point is that this notion of conditional reliability can be employed by the reliabilist to account for the positive epistemic status of some beliefs that are not justified. If

(categorical) reliability yields (categorical) justification, we can say that conditional reliability yields “conditional justification”: the output belief is/would be justified if the input beliefs are/were. Conditional justification is not as good as categorical justification, but it is an epistemic good. This is quite plausible even independent of the present concerns. Forgetting reliabilism for a moment, suppose an agent draws an impeccable inference from an unjustified belief. The resulting belief is unjustified, but to say only this is to ignore the fact that the agent has accomplished some positive epistemic achievement by drawing a proper inference, despite the flawed premise. An agent who draws valid inferences from unjustified premises is doing something epistemically right, something that an agent who draws invalid inferences is not. Being conditionally justified in this way is a positive epistemic accomplishment, and having conditional justification is a positive epistemic status, even if subordinate to being justified. Being justified is a positive epistemic status that is nevertheless subordinate to knowing; why not allow that conditional justification is subordinate to justification though still positive?

The reliabilist, therefore, can claim that some of the demonworlders’ beliefs have positive epistemic status, even though they are not justified. This meets, or at least goes a long way toward meeting, the intraworld challenge to discern a difference between agents who are using different reasoning processes: the processes might be equally unreliable but still differ in conditional reliability. It doesn’t meet the interworld challenge to discern *no* epistemic difference between us and the demonworlders. Conditional justification is not the same as justification, so one who demands that the demonworlders are just as justified as we are will be unsatisfied.

I want to say three things about why the gap between justification and conditional justification may not be as great as it seems. First, it is good to keep in mind that when a

reliabilist claims that a belief is unjustified, there is no imputation of irrationality or the like involved. An agent might be unjustified in a belief without having any reason to think that the belief is false, or that it is unjustified; she might not even be in a position to have any such reason; she might even be justified in thinking of the unjustified belief that it is justified. On some epistemologies, to say that *S* is unjustified is to imply, at least suggest, that the agent ought to know better than to hold that belief. This is clearly not true for reliabilism, and especially when the justification failure is due to demonic intervention.⁶ So there is no tension between the claim that a belief is unjustified and the claim that it has a positive epistemic status.

Second, as sketched above, the ‘unjustified’ verdict for conditionally reliable but categorically unreliable processes results not from the reliabilism per se, but from the requirement that the inputs to belief-dependent processes themselves be justified.⁷ One could imagine an internalist who endorsed a conception of justification so subjective and so synchronic that it allowed cogent inference from unjustified premises to yield justification. Disagreement with this view should not be confused with disagreement over what counts as a good inference. If the justified-input restriction on belief-dependent processes is the only reason for the downgrade from ‘justified’ to ‘conditionally justified,’ the complaint that reliabilism has failed to preserve interworld epistemic identity is misplaced: when the reliabilist and the internalist disagree about whether a belief is justified or merely conditionally justified, the disagreement is not about the justification-preserving powers of memory, or deduction, or the like. The disagreement is about something else entirely.

⁶ To say that it is not irrational is not to say that it’s “weakly justified” (Goldman 1988) or “personally justified” (Engel 1992) or the like.

⁷ I mean psychologically belief-dependent here. Where the context disambiguates, I won’t bother with the qualifier.

Third, the insistence that we discern no difference is not only misplaced, but quite resistible. The reliabilist who imposes a justified-input requirement for memory is not therefore required to deny that reliably remembered beliefs with unjustified originals have no positive epistemic status. Remembering reliably, like deducing correctly, is an epistemic accomplishment, and the reliabilist is under no obligation to deny this simply because she claims that the output belief is unjustified. It might yet be conditionally justified, and the reliabilist might invoke this positive epistemic status to explain away the intuitions of those who think that memory is a generative source of justification: the generationist is right to think memory beliefs with unjustified originals have positive epistemic status, but is wrong to mistake that positive epistemic status for justification. Conditional justification is not only a positive epistemic status, one might hold, it's so close to justification that generationists have confused the one for the other.

In any case, there is *something* to be said in favor of the memory beliefs of demonworlders or normal humans who correctly remember unjustified beliefs---they are conditionally justified---and this ought to be good enough for the reliabilist. If the reliabilist can make good on the claim that the perceptual input beliefs are unjustified, then the merely conditional nature of the memory justification should be quite good enough.

I want to put these three points together and claim that the test cases for reliabilism, vis-a-vis NDP, anyhow, are cases of belief-independent belief. Conditional justification gives the reliabilist something conciliatory enough to say regarding belief-dependent beliefs.

The consequences are more significant than might at first appear. Suppose instead of the psychologized Cartesian story, one offered something like the following:

S has the standing belief that the best explanation for her having consistent and uniform experiences is the overall reliability of the senses (that is, that things are generally how they appear). From this and *S*'s introspective belief that she's appeared to redly, *S* concludes that there's (probably) something red nearby.

I think there are good reasons to think that perceptual beliefs are epistemologically basic and hence belief-independent, and I think the relevant arguments are more epistemological than psychological in nature (i.e., it's because the justification of the perceptual beliefs does not depend on the justification of the other beliefs) (Lyons 2009a). But I don't want to insist on any of that now; let's take seriously this nonCartesian proposal in favor of the belief-dependence of perception.

The beliefs just cited include one that results from introspection, one that results from abduction, and one that results from statistical inference (in particular, direct inference). Statistical inference is as conditionally reliable in demon worlds as in the actual world (that is, in demon worlds that are uniform enough as to be indistinguishable from the actual world),⁸ so if abduction is also as conditionally reliable, we can treat this in just the same way we treated the Cartesian proposal earlier, and the reliabilist holds that the belief is justified even though the overall process is unreliable, because the component processes were either categorically or conditionally reliable, as the theory demanded.

⁸ Are direct inference and other forms of inductive inference conditionally reliable in phenomenologically consistent demon worlds? One way to understand conditional reliability is in terms of actual frequencies, ignoring those instances where falsehoods are/have been among the inputs. Alternatively, one might use a counterfactual approach: the process would produce a large number of truths were it to receive true inputs. Neither way seems to pose an insurmountable obstacle for the claim that these types of inference are conditionally reliable in worlds governed by consistent demons.

Suppose, though, that abduction is not conditionally reliable in the demon world.⁹ Even then the reliabilist can claim that the resulting belief, though unjustified, is conditionally justified, and this should be concession enough. Granted, demonworlders' perceptual beliefs don't have the same epistemic status as ours, but they do have positive epistemic status, and reliabilism can account for this with its original machinery of belief-dependent and belief-independent processes. So for belief-dependent beliefs, reliabilism is well-positioned to give the internalists enough of what they want. All the more so if I can show (in the next section) that it really is plausible to deny that the belief-independent beliefs are justified and can trace the gap between full justification and mere conditional justification to these.

Of course, internalists who are attracted to the kind of story just offered propose it as an epistemological account of perception, not a psychological account. This is an unstable position, however. What the reliabilist is clearly after is a theory of *ex post*, doxastic justification, what Feldman and Conee (1985) call "well-foundedness": the agent must be believing for the right reasons; mere possession of such reasons is not enough. Given this concern, an appeal to the agent's belief in the reliability of perception can't really be getting at part of the agent's justification for holding some perceptual belief, unless the perceptual belief is in some sense causally or counterfactually dependent on the reliability belief, at least in the sense that if the agent didn't have the reliability belief, she wouldn't have or wouldn't retain the perceptual belief. But this is sufficient for a belief to be one of the inputs to a cognitive process. There is no reason

⁹ I'm frankly not sure whether abduction is or is not conditionally reliable in demon worlds, and a careful examination of this question is beyond the scope of this paper. Deciding on this will depend on the psychological details. Do we rationally intuit that theory *A* is simpler than theory *B*, and that simpler theories are more likely to be true, and then deduce that *A* is likelier than *B*? If so, abduction will parallel the Cartesian account of perception discussed above, and abductive inference is justification conferring in demon worlds. To argue that reliabilism gets the wrong verdict regarding abduction, one would have to argue in some detail that abduction is not conditionally reliable in demon worlds. To do that, one would have to offer a different account of abduction than the one just sketched. I don't take the matter of abduction in demon worlds to be settled, but I'll leave it as unfinished business.

reliabilism should eschew tacit beliefs as inputs, so long as they make a difference as to what outputs are produced.¹⁰ As an illustration, recall the Cartesian theory. Even if it succeeded on other grounds, it would fail as an account of what *in fact* justifies *our* perceptual beliefs if it turned out that our perceptual beliefs were insensitive to our religious beliefs (as they almost certainly are).

So anyone who holds that our perceptual beliefs epistemically depend on other beliefs is subject to the following dilemma. Either the epistemological account has sufficient psychological reality to underwrite a theory of ex post, doxastic justification, or it doesn't. If it doesn't, then it only tells us how agents *could be* doxastically justified in their perceptual beliefs, not how they *are*; it indicates neither how nor whether our perceptual beliefs really are ex post, doxastically justified. If the account does have psychological reality, then it's the reliabilist theory of belief-dependent, rather than belief-independent beliefs that's at issue, but that theory doesn't give the offending verdict about demon worlds. So once again, if one is going to press a NDP attack on reliabilism, it really has to focus on belief-independent beliefs, that is, beliefs that are both psychologically and epistemologically belief-independent.

To summarize, with the possible exception of abduction (see note 9 above), it's really just perceptual beliefs that reliabilism gets the obviously "wrong" answer on and even then, only because the reliabilist assumes that perception is belief-independent. To pursue NDP, then, let's

¹⁰ Whether or not reliabilism is right to hold that all psychologically belief-dependent beliefs are thereby epistemologically belief-dependent, the converse is true: all epistemologically belief-dependent beliefs are psychologically belief-dependent. If it helps, we can think of such background beliefs as being encoded not as activation patterns across the input layer of a connectionist network, but as being encoded in the weights. Such states aren't beliefs in the transient and occurrent sense of 'belief', but this is not the only legitimate sense of 'belief', even for a reliabilist. Beliefs of this sort can make a causal difference, in that without them, different output patterns would result, and these beliefs also make (or so we are supposing) an epistemic difference, in that if the agent were not justified in these standing, background beliefs, she would not be justified in the output beliefs. Together, these seem to me sufficient reason to treat the background beliefs as inputs to a psychologically belief-dependent process.

embrace the independently plausible view that perception is belief-independent, and try to suppose that my demon world counterpart has the same justification for his beliefs as I have for mine. Together, these lead to a view I call “Seemings Internalism”:

(SI): If S is appeared to as if p , then S is *prima facie* justified in believing that p .

Is this view really as plausible as its popularity would suggest?¹¹ In particular, does it have any advantage over reliabilism in its verdicts about demonworlders?

3. Belief-independent beliefs in a demon world

I have tried to show that if NDP is to cause any real embarrassment to reliabilists, it will have to be because reliabilism is wrong to insist that the *belief-independent beliefs* of demonworlders are unjustified. Since my attention is restricted to evidential versions of NDP, this comes down to the question of whether the evidence we have for our belief-independent beliefs (in particular, our perceptual beliefs) mandates that our demon world counterparts believe as we do. The question, that is, is whether the belief-independent evidence for our perceptual beliefs---our experience---requires demonworlders to adopt our perceptual beliefs as well.

I use ‘mandate’ and ‘requires’ intentionally here, for it is important to see what the proponent of NDP is presupposing. Clever presentations of NDP encourage the reader to view the reliabilist as imposing draconian strictures on justification. It seems as if the reliabilist “blames the victim,” forbidding her from doing what we get to do, simply because she has the bad fortune to be in a demon world. By contrast, the internalist is seen as being far more

¹¹ This view has been held by a great number of epistemologists, starting, perhaps, with Reid ([1785] 1967). Contemporary and explicit defenses of the view begin, I think, with Quinton (1966), though Pollock (1974, 1986, 1995; Pollock and Cruz 1999) probably deserves the most credit for really developing and articulating the view. Part of its current popularity may be due to proponents like Huemer (2001) and Pryor (2000), who call it “Phenomenal Conservatism” and “Dogmatism,” respectively.

generous, lenient, and forgiving about belief formation. But in truth, the internalist is not more lenient. To say that my demon world counterpart is justified in exactly those beliefs I would be justified in believing is to prohibit him from forming beliefs in certain ways, and that prohibition is not well motivated.

Consider the following little story:

The travelers passing the time at Grand Interworld Station are a diverse lot. There are terrestrial humans of all eras, but there are also centaurs, unicorns, Norse deities, sentient teapots, talking lions, swampmen, inhabitants of Mars, Alpha Centauri, and Twin Earth, and more. Often they ignore each other and wait for a train back to their own possible world. But today a number of them are gathered around listening to a Cartesian demon discuss its ongoing exploits with poor Larry. The demon directly produces in Larry various sensations, E_1 , E_2 , etc., to which Larry responds by believing, respectively, that there's something red nearby, that there's something wet, etc. The demon is, of course, able to produce the relevant sensations in the minds of its listeners as a means of telling its story, which makes the demon quite a popular storyteller.

The demon is expressing dismay, however, because Larry is getting everything systematically wrong. He responds to E_1 with the belief that something is red, when he's supposed to respond by believing that something is green; E_2 was supposed to produce the belief that there's something soft, but Larry responds by believing there's something wet. The humans listening to the story, like the demon, are puzzled by Larry's apparent irrationality. Something that looks a bit like a toilet plunger speaks up: "I don't see what's the matter. Larry is perfectly justified in his beliefs. I don't know how things are on Earth or wherever you demons are from, but for us on Tralfamadore, what you're calling E_1 is the sensation we associate with redness; what you're calling E_2 is the one we associate with wetness, and so on. Larry is believing just as we would, and just as he should."

After making sure that the demon is, in fact, producing the same sensations in all its listeners, it becomes clear that the audience can't agree about Larry's epistemic status. The colortemps (Markie 2004) are baffled that someone would believe that something is red or green on the basis of a sensation that clearly indicates that an object is 78 degrees F. The demon sometimes produces religious experiences in Larry, which he takes as evidence for the existence of a loving and powerful God. Plantinga and Alston think he's right to do so; W. K. Clifford think he's being gullible. There are also disagreements about defeaters. Larry sometimes feels a vague sense of dread, but he doesn't take it seriously; if inductive evidence indicates the situation is going to be okay, he goes with induction. Most of the humans approve of this; Spiderman and Socrates think he's being irrational.¹² He sometimes has visual imagery of famous people pop into his head, but if

¹² Spiderman is said to have a "spider sense," which warns him of impending danger; Socrates was said (by himself, apparently) to have had a daemon that did much the same for him.

he happens to catch an image of the president in front of a notable New York landmark, he persists in believing the news report that says the president is in Washington. The clairvoyants in the audience disapprove of his disregarding important defeaters.

Unsurprisingly, many of the audience members respond to their experiences in a different way than do other audience members. Equally unsurprisingly, each demands that Larry respond to his experiences in the way they would.

This fable suggests that our intuitive judgments about demonworlders are chauvinistic and parochial, and therefore not to be taken seriously. We think Larry is justified when he's forming beliefs the way we should, but if we were from Tralfamadore, we'd want Larry to form beliefs in a very different way. The problem is that we can't allow Larry to form beliefs in (just) our preferred manner without prohibiting him from forming them in some other manner, and there's no good reason for this prohibition.

Let us say that a *cognitive mapping* is a function from experiences to beliefs. An agent satisfies the cognitive mapping $f(e) = b$ just in case whenever that agent has experience e it forms belief b . An *epistemic mapping* is an assignment of justificational status (justified, unjustified) to experience/belief mappings. An agent satisfies the epistemic mapping $g(e, f(e)) = j$ just in case the value of j is 'justified' whenever e justifies that agent in believing $f(e)$. Different cognizers will satisfy different epistemic mappings. Some of this will be uninteresting and due only to the fact that they have different experiential repertoires and therefore cognitive mappings with different domains. But some will be due to the fact that different cognizers are justified in different, incompatible, experience-belief mappings. If we represent epistemic mappings as sets of experience-belief-justificational status triples, we can say that two epistemic mappings *conflict* iff they contain one or more triples that differ only in the third position. Two agents therefore will satisfy conflicting epistemic mappings if the same experience makes the one but not the other justified in some belief.

The Grand Interworld Station story is supposed to make plausible the claim that it is possible for two agents to satisfy conflicting epistemic mappings.¹³ The Tralfamadorians' epistemic mapping conflicts with ours (experience-belief mappings that are justified for us are unjustified for them, and conversely); but the Tralfamadorians are no less justified for that. If this plausible claim is true, then what are we to do about poor Larry? "Allow" him to satisfy our cognitive mapping, in effect requiring him to do so, and prohibiting him from satisfying other, conflicting, cognitive mappings? This seems arbitrary and chauvinistic. Better, it seems, to prohibit all experience-belief mappings on the grounds that they are equally unreliable than to endorse one over the others on the grounds that it is *ours*.¹⁴

Here's a convergent line of thinking: we would offer different verdicts about the demonworlders if we came from different worlds or had different cognitive capacities, and this fact suggests that our verdicts about demonworlders are not so much judgments about what *they* should believe but about what *we* should believe were we to undergo the experiences they're undergoing. But this latter judgment is no threat to reliabilism. If *I* were undergoing those experiences, I'd be justified in the resulting beliefs, true; but if I were undergoing those experiences, those beliefs would be reliably formed, so that fact is no threat to reliabilism. It is only if *they*, the demonworlders, are justified in their unreliably formed beliefs that reliabilism is in any danger.

¹³ This assumes that there are some belief-independent beliefs. Epistemic mappings can't conflict unless one counts a certain experience-belief mapping as justified while another counts it as unjustified, and this can't happen unless some experience-belief mappings are justified, and this can't happen unless some experiences justify independently of other beliefs, i.e., unless there are belief-independent beliefs. The existence of belief-independent beliefs is the part of the story that SI gets right.

¹⁴ One is reminded of Richard Dawkins's recent claims to the effect that we are *all* atheists about the vast majority of the gods. The difference between Christians and Dawkins is that he goes just one god further.

I have been taking the notions of a demon world counterpart, and an epistemic counterpart more generally, as primitive. What features must an agent share to be justificationaly identical to me? Obviously, if he shares all my properties, including the relational properties, he and I will be equally justified in all our beliefs, but just as obviously, this is no worry for reliabilism. The intuitive answer, and the answer that serves as the starting point for NDP, is that any two experientially and doxastically identical agents are justificationaly identical as well, so x and y will count as epistemic counterparts just if they are experientially and doxastically identical.

The problems we have been seeing here, however, result directly from our thinking of the demonworlders only in terms of their beliefs and experiences. NDP presupposes a reductionistic view of epistemic agents as nothing more than cognitive mappings. Back to the Grand Interworld Station:

At some point, someone hits on a way to resolve the disputes. They ask whether Larry is a human, a Tralfamadorian, or what. Each of the listeners had been assuming Larry was one of their own. (It turns out ‘Larry’ is a *very* common name.) The question seems to get at something important, for if Larry is a normal human, then perhaps certain experience-belief mappings are appropriate; otherwise, different, conflicting, experience-belief mappings are appropriate, depending on just who and what Larry is. The demon’s response is unfortunate: “Oh, Larry isn’t any of those things; he’s a disembodied spirit I manufactured by breathing on a lump of clay.”

The idea here is that if the demonworlder could somehow be “anchored” to some type of world, or body, or environment, or cognitive architecture, or phylogenetic history, or *something*, we could make a nonarbitrary choice among conflicting epistemic mappings. That is, if Larry somehow *belongs* in some environment, or possible world, or the like, then this might bind him to some particular cognitive mapping or mappings, thus ruling out the others. I think this is very plausible, but it is off the table as far as NDP is concerned, for NDP is supposed to show that *any*

creature with the same experiences as I have is therefore justified in forming the same perceptual beliefs as I am, that our psychological counterparts are our epistemic counterparts as well, and this is supposed to show that justification is determined by the agent's psychological states---in the case of belief-independent beliefs, by the agent's experiential states. But then the right epistemic mapping must be the same for everyone; "unanchored" demonworlders should be no different from the "anchored" ones. But this seems wrong: if Larry is really unanchored, then he has no right to arbitrarily "choose" one cognitive mapping over the others; there is nothing to make the cognitive mapping he satisfies *better* than the ones he does not satisfy.

Lehrer and Cohen may be alive to this; perhaps this is why their version of NDP, quoted above, features *us* as the victims of the demon. NDP is supposed to argue for, or at least pump the intuition in favor of, mentalism: the view that if two agents are alike with respect to all their nonfactive mental states, then they are alike with respect to justification. But we *are* anchored to a world/environment/etc., so if the plausibility of the NDP argument for mentalism derives from it being *us* who are deceived by the demon, that plausibility is spurious. The Interworld Station story, by focussing on agents who are our experiential/doxastic counterparts but not in any sense *us*, is supposed to pump the opposite intuition: there are conflicting epistemic mappings, which shows that justification is not determined entirely by mental states, that is, not entirely by cognitive mappings, but also by anchoring---to world, body, species, etc.

Recall the normal worlds solution to NDP and the problems with it. According to normal worlds reliabilism (Goldman 1986), justification is a matter not of reliability in situ, but in *normal worlds*, i.e., worlds that conform to our general beliefs about how the actual world is. Normal worlds reliabilism answered not only NDP but the famous clairvoyance objections

(BonJour 1985) to in situ reliabilism. Clairvoyance is not reliable in normal worlds, so clairvoyance is not a justification-conferring cognitive process. One problem with the normal worlds view, however, was just this: it implied that clairvoyance *couldn't* yield justified beliefs---for any kind of creature---and this seems too restrictive. Treating unanchored demonworlders as if they were anchored to our world poses a very similar problem. Somewhere in the space of possible worlds there are creatures with reliable clairvoyant powers, who are justified in their clairvoyance beliefs, and whose experiential evidence is just like something that would not give us evidence of distant facts. When I have a mind's-eye visual image of the president in New York, I'm not (prima facie) justified; some possible cognizer, however, is.

The deeper problem for normal worlds reliabilism, however, was that it seemed arbitrary, parochial, and chauvinistic to deny justification to others, simply because their worlds aren't the way we imagine ours to be. It would be a more generous but still illicit sort of favoritism to grant justification to their beliefs, which are inappropriate in their world, but would be appropriate if their world were the way we think ours is. It is similarly unprincipled and provincial to avow justified (/unjustified) status for the beliefs of unanchored demonworlders, simply because those beliefs would be justified (/unjustified) if those agents were anchored *here*. Man is not the measure of all things.

Sosa (2007) and others have held that justified belief is a matter of exercising a certain kind of skill. Let's pursue the analogy between perceptual belief and motor skills a bit in the context of demon worlds. It is interesting in this connection to ask what skills, if any, Larry possesses. Is he, for instance, a good swimmer or a good violinist? In some uninteresting sense Larry is neither, since he doesn't have arms with which to swim or operate a violin. But this

quick answer masks a more interesting question about what it takes to have a skill. A disembodied brain can still have motor skills, in the sense of having a kind of procedural knowledge: motor programs and schemas that would allow the brain, if properly embodied, to swim well or play the violin well. Could such motor programs make an unanchored demonworlder a good violinist or a good swimmer? I think not, for the sorts of reasons we have been seeing.

Michael Phelps is superb at swimming in his body, in liquids of a certain viscosity and density, in gravitational fields of a certain strength, with certain laws of nature operative. But surely his swimming ability is not robustly resistant to perturbation by changes in these variables. And clearly, this is no real knock on his virtues as a swimmer.¹⁵ He is little or no less skilled as a swimmer for not having mastered the art of propelling his body through motor oil (a task which, one might assume, would require very different strokes and strategies) or of moving through water in a young girl's body---or a dog's, or a fish's---instead of his own. Does a disembodied and unanchored psychological duplicate of Phelps count as a skilled swimmer merely because there is some set of possible body/environment/world complexes in which he is actually good at swimming? By that token I would be---rather, I *am*---a talented violinist: I have enough motor control over my vocal organs that, were the peripheral nerves rewired appropriately, I'd be able, maybe after a few minutes of practice, to use my speech routines to control a violin with a great deal of skill. (Well, maybe not a violin, but probably a banjo, and certainly a bagpipe; nevertheless, I am not a talented bagpiper.) The fact that my motor schemas

¹⁵ Though see Henderson and Horgan (2006) for a contrary view, at least for the epistemological analogue.

would lead to superb performance in some possible worlds, though not in the world I'm in, indicates perhaps that I could in some sense be skilled at x , not that I *am* skilled.¹⁶

I have been arguing that different possible agents satisfy different cognitive mappings, and on the face of it, at least, they are right to do so; i.e., they satisfy different and conflicting epistemic mappings. If so, then for our unanchored demonworlders, we must either choose one cognitive mapping from the (infinitely) many and allow/require them to satisfy it, or deny that the satisfaction of any cognitive mapping makes their beliefs justified.¹⁷ The first option is unprincipled and chauvinistic; this is why I have been recommending the second.

This is all conditional, however, on the claim that possible cognizers satisfy epistemic mappings that conflict with ours. There are two credible ways in which one might try to deny this, without giving up on Seemings Internalism and the view that perception is belief-independent. I turn to these next.

4. Evidence Essentialism

“Evidence Essentialism” is the view that evidential relations hold necessarily:

(EE): If e is evidence of h for S , then necessarily, and for any S , e is evidence of h .

It is an intuitively plausible view, provided we read ‘ e ’ as picking out the whole of S ’s evidence regarding h , not just some part of it. Evidence Essentialism, however, would imply that factors

¹⁶ None of this, of course, is to claim that the unanchored Phelps duplicate is a *bad* swimmer either (on the grounds that there’s a range of possible body/environment/world complexes that he swims poorly in); that would be to make the same mistake in the other direction. As mentioned briefly earlier, when the reliabilist claims a belief is unjustified, there is no implicature that the agent is thereby doing anything *wrong*, just that she’s not doing anything right.

¹⁷ Could we just allow them to be justified in whatever cognitive mapping they satisfy? This would be incompatible with the mentalist thesis that NDP is supposed to establish (since some demonworlders will be forming beliefs in ways that are identical to their actual counterparts).

like the agent's environment or anchoring, contingent probabilistic relations between e and h , and psychological factors that are not part of the agent's evidence (e.g., cognitive capacities, learning history, etc.), are irrelevant to justification.

Seemings Internalism (SI) is committed to Evidence Essentialism (EE), at least as far as seemings are concerned, but one could hold EE without endorsing SI. SI results from combining EE with evidentialism (the view that justification is entirely determined by one's evidence) and the claim that perceptual beliefs (or whatever beliefs have corresponding seemings states) are belief-independent.¹⁸ The Grand Interworld Station story is supposed to show that there are conflicting, satisfied, epistemic mappings, and thus that SI is false. If, as we have been assuming, perception is belief-independent, and evidentialism is true, then EE is false as well.

There are two ways one might try to retain EE and thereby SI, in an effort to maintain the plausibility and nonarbitrariness of the claim that unanchored demonworlders are justified in some (but not all) of their beliefs. The first is to insist that seemings, or experiential states, are more abstract than I have been assuming; the second is to endorse something like Pollock's (1986) epistemological theory of concepts.

One way to retain Evidence Essentialism (not to argue for it, but to resist the Interworld Station objection to it) is to claim that I have misidentified the relevant elements of the cognitive mapping: the nondoxastic experiences that do the epistemic work are high level and abstract states, not the states that differ among the passengers at the Interworld Station. Take the case of color inversion without error: you and I both believe there's something red, even though the way it looks to you is the spectral opposite of the way it looks to me, and so on with the other colors.

¹⁸ One could endorse EE but not SI by rejecting evidentialism (e.g., hold that basic beliefs aren't based on evidence at all) or by requiring that basic beliefs be about the agent's mental states, not about the external world.

The same low level sensation might accompany my justified belief that it's red and your justified belief that it's green, thus apparently vitiating EE and SI. The current proposal, however, is that we interpose an abstract experiential state between the sensations (which differ) and the beliefs (which do not). If the high level, abstract, appearances are the same, and these appearances are the experiential states invoked by SI to serve as evidence for perceptual beliefs, then both EE and SI can be retained. Even though you and I have different sensations, it *looks red* to both of us, in the relevant sense of 'looks'. This "looking red" is a nondoxastic experiential state distinct from the raw sensory phenomenology and more like the "seeming true" that attaches to certain a priori propositions. If cognitive and epistemic mappings are concerned with these abstract appearances, there need not be any conflict among epistemic mappings.

Such a view might also explain how something can look---in some sense---the same to the expert and the novice, even though they are justified in different perceptual beliefs. The bird produces the same sensation in the expert and the novice, but the expert is justified in believing it's a pileated woodpecker, while the novice is not (Cf. Lyons 2005, Feldman 2003). Perhaps this is because the expert is appeared to pileated-woodpecker-ly while the novice is only appeared to birdly. That is, the expert has a high level appearance with the content *pileated woodpecker*, and the novice does not.

Notice, however, how far these high level, abstract appearances are from the sensory states we normally associate with perceptual experience. There is nothing particularly spatial about my pileated woodpecker appearance, for example, and there is nothing visually *red* about the invert's being appeared to redly. Instead, the appearance is more like a seeming-true, or an attraction to believe.

I've addressed this in a few other places (Lyons 2005, 2009a, 2009b), where I argue that in cases where the agent does believe that p , such seeming as if p is not distinct from that belief and therefore not something that can serve to confer evidence on that belief. I won't rehash that here. Let's consider instead the proposal that p 's seeming true is not a distinct *state* from the agent's believing that p (in the way the sensory experience would be), but rather, a phenomenal *property* of the agent's believing that p , in virtue of which that believing is justified.¹⁹ The problem with this, I think, is that if seeming true thus construed could confer justification on any beliefs, all (nonvoluntary) beliefs would be justified, for all nonvoluntary beliefs seem true, in this sense, to their possessors. But of course not all beliefs are justified, not even *prima facie* justified, so such seeming true doesn't confer justification.

It would be natural to reply by insisting that not all beliefs have this seeming-true property. The belief that $2+2=4$ does, but the Banach-Tarski theorem, for instance, doesn't; in fact the Banach-Tarski theorem seems false. Now, certainly there's a phenomenology of obviousness that is connected to the consideration of simple *a priori* truths, which is not connected to the consideration of other truths. I can't so much as consider my own existence without its seeming true, though I can consider your existence without its seeming true. But this doesn't show that this seeming true, this "truthiness," isn't a property of all *beliefs*; it only shows that it's not a property that attaches to belief contents, irrespective of the propositional attitude taken toward that content. My claim was not that "truthiness" attached to the content, but rather

¹⁹ It would be nice to name names, and I would if I could. I *think* this is the sort of thing Michael Huemer (2001) would endorse, but I can't be sure what any of the proponents of Seemings Internalism really understand their seemings to be. I have tried in other places (Lyons 2009a, 2009b, 2010) to point out the need for clarification on this point.

to the attitude of believing. Thus, I can't believe something that doesn't seem true, even though I believe many things now that didn't always seem true, and might not seem true later.

Though I can consider the Banach-Tarski theorem without its seeming true, I can't (nonvoluntarily) believe it without its seeming true. The fact that logicians I respect tell me it's true makes it seem true to me. This is compatible with its seeming false, just as in the waterfall illusion, things seem to move and to stand still at the same time.

It is possible in this way to understand the claim that the same thing "looks red" to both the normal and the invert, even though their sensations are radically different. But the claim is epistemically irrelevant, for "looking red" in this sense is little or nothing more than seeming true. If there's yet another kind of "looking red" beyond the (perceptual) belief that something is red, beyond the phenomenal "truthiness" attached to that belief, and beyond the sensory experience, it would be good to know what it might be.

Thus, I don't think that the appeal to high level, abstract appearances makes for a very plausible epistemology. Consequently, it does little to defend SI and EE against the Interworld Station objection. In addition, recall that our central question is whether NDP poses a crippling objection to in situ reliabilism. SI, especially when augmented with the high level, abstract understanding of appearances as just detailed, is a controversial view, one that is not at all obviously true. Surely not even all the proponents of NDP hold this odd view about nondoxastic experience. If it is only on the prior assumption of this controversial form of Seemings Internalism that NDP threatens reliabilism, this is good news for reliabilism.

Pollock's epistemological theory of concepts offers a second way to defend EE. The theory is roughly this: included in the content of any nonlogical concept is a specification of the

criteria that would count as evidence for the instantiation of that concept. Thus, if I take E_1 as evidence of redness and you don't, then you and I are not operating with the same concept of *red*. This implies that the Interworld Station myth is confused. If the Tralfamadorians disagree with us about what is evidence of wetness, then we and the Tralfamadorians don't share a concept of *wet*. But that means there couldn't be conflicting epistemic mappings after all; there couldn't even be conflicting cognitive mappings.

Obviously, I will deny the epistemological theory of concepts. It strikes me that the fact that the theory implies that the Tralfamadorians couldn't have the same concept of wetness---and thus couldn't share my belief that water is wet---is already more evidence against that theory than has been adduced in favor of it. Furthermore, since I, like some other epistemologists, think that nondoxastic experiences don't serve as reasons, or evidence, it would seem to follow that I couldn't agree with Pollock about what color the chair is, whether there even is a chair, and so on.

A decisive disproof of the epistemological theory of concepts would require more than this and would take us quite far afield. But a decisive disproof is not required for the present purposes. NDP is supposed to offer a straightforward refutation of reliabilism, a pithy counterexample that has intuitive pull on everyone and can be embraced and wielded by any internalist. But the epistemological theory of concepts is far from obvious. If it is indispensable to the NDP attack on reliabilism, the argument comes down to the conclusion that we cannot embrace both reliabilism and the epistemological theory of concepts. This is hardly just a roundabout way of stating that reliabilism is false. The reliabilist can reject the theory of concepts without undue distress. One could, of course, develop a positive argument to the effect

that the epistemological theory of concepts is true; therefore, Evidence Essentialism is true; therefore in situ reliabilism is false. By then, however, NDP has fallen out as irrelevant.

5. The New Evil Demon Without Seemings Internalism

Seemings Internalism founders on the problem of unanchored demonworlders, developed in section 3. The attempts to salvage SI, examined in section 4, did not look promising. Let us abandon SI, then, and see what remains of NDP.

SI holds that if two cognizers are experientially identical, they are justificationally identical. Mentalism (Conee and Feldman 2004) holds that if two cognizers are experientially identical, they are justificationally identical. The latter appeals to a broader supervenience base and so is clearly more likely to be true. One easy way to relax the restrictions of SI is to suppose that the agent's other beliefs contribute to the justification of her perceptual beliefs, that experiences justify perceptual beliefs only in conjunction with other, background beliefs. We would want to replace the understanding of cognitive mappings developed earlier with one that construes them as functions from conjunctions of experiences and background beliefs to perceptual beliefs. We would reconceive epistemic mappings accordingly. One might then argue that there aren't conflicting epistemic mappings after all, so long as the different possible perceivers have different background beliefs about which experiences reliably indicate which external states of affairs.

There are well known challenges for such a view and correspondingly good reasons for thinking that perception is belief-independent (I would surmise that the belief-independence of perception is the predominant view, even among internalists). I will pass these by in favor of the

dilemma rehearsed above: if the perceptual belief depends causally or counterfactually on the background beliefs, then the reliabilist can claim that the perceptual beliefs are belief-dependent; if not, then the theory does not address the question of doxastic, ex post, justification. We would need to know the details about these background beliefs and how they figure into the justification of perceptual beliefs, but the reliabilist will be able to make some sort of appeal to conditional justification. How mollifying this move will be will depend largely on the details of the psychological story offered by the internalist.

We were driven to SI by the role of mentalism in NDP, along with the need for NDP to focus on belief-independent beliefs. How can mentalism render perception belief-independent without embracing SI? It is possible to appeal to background beliefs, ascribing to them a metaevidential, rather than an evidential, role. That is, the background beliefs, instead of serving as part of the agent's evidence for the perceptual belief, might play the role of determining what counts as evidence for what. Richard Foley (1987) holds that *e* is evidence for *S* of *h* just in case *S* tacitly believes that *e* renders *h* probable, i.e., if *S* were to be sufficiently reflective, she would believe that *e* renders *h* sufficiently probable. I am presuming that this higher order belief is not posited as an additional premise in *S*'s argument for *h* but rather, an enabling condition that makes *e* adequate evidence of *h* (for *S*). Peter Markie (2004) holds a similar view, which he couches in terms of internalized epistemic norms and articulates explicitly in terms of bridging the gap between nondoxastic experiences and perceptual beliefs.²⁰ Granting the background beliefs a metaevidential---hence nonevidential---status would block my conditional reliability move; they aren't evidence, thus aren't premises, thus don't render the process belief-dependent.

²⁰ I am not sure how Markie conceives of the norms and thus not sure whether his view should be classified here as one that appeals to metaevidential *beliefs*, or whether it should be classified with the next group of internalist theories I discuss.

They don't even need to be justified in order to fulfill their epistemic role. This move would allow us to revert to the original cognitive mappings involving just experiences and perceptual beliefs, while admitting the existence of conflicting epistemic mappings. Unanchored demonworlders would have metaevidential background beliefs and therefore be justified in some but not all of their belief-independent beliefs.

There's much to be said in favor of this sort of view, provided that one can swallow a subjectivist epistemology. Most of us cannot; we think that evidential relations are objective, at least in the sense that your sincerely believing that affirming the consequent is valid---or your having internalized the norm of affirming the consequent---does little or nothing to make the conjunction of q and $p \supset q$ good evidence for p , especially if, e.g., you never had any decent excuse for believing that affirming the consequent was valid but have always endorsed it to spite your logic teacher.

As before, this would be too quick for a conclusive dismissal of subjectivism. But all I'm trying to do right here is to show that reliabilism has plausible responses to the various twists on NDP.

There is one last internalist view that needs to be considered. Mentalism, again, is the view that if two agents are mentally identical, then they are justificationaly identical. I have been taking mentalism to be a kind of "constrained mentalism," according to which the only relevant mental states or properties are the currently available beliefs and experiences of the cognizer. But what counts or should count as "mental" is notoriously unclear, and perhaps the mentalist can get some mileage out of this fact. A "liberated mentalism" would claim that

justification supervenes on the mental, in some larger sense of ‘mental’.²¹ A liberated mentalist could, for example, appeal to past experiences not currently remembered, or cognitive architectural facts (e.g., that the belief was produced by a certain kind of module), or factive mental states of the agent, and so on. Not all of these mental properties are even possibly shared by our demon world counterparts, but some are. Such properties might allow the liberated mentalist to admit the possibility of conflicting epistemic mappings, while still insisting that my demon world mental duplicate is justified in the same beliefs I am.

To illustrate, let’s consider a liberated mentalist theory I think is very nearly true. Suppose that the novice and the ornithologist look out, have identical experiences, and both come to believe that there’s a pileated woodpecker in front of them. According to a “perceptual systems theory” of perception (Lyons 2005, 2009a), the main difference between them is that the expert’s pileated woodpecker belief is the output of one of her perceptual modules, while the novice’s belief comes from somewhere else. They satisfy the same experience/belief mapping, but because of the cognitive architectural---*mental*---differences, one is justified and the other is not. I’m not sure whether my disembodied counterparts would have cognitive modules or not, but my molecule-for-molecule brain-in-a-vat duplicate certainly would; he would have whatever perceptual expertise I do. The perceptual systems theory could thus capture the claim that my duplicate is justified in just those cognitive mappings I am justified in.

A more popular mentalist view holds that the agent’s actual history might matter to justification; that is, her training or learning history might affect which experiences justify which

²¹ I use ‘constrained’ and ‘liberated’ because ‘narrow’ and ‘broad’ would misleadingly suggest the semantic issues set aside at the beginning of the paper.

perceptual beliefs. And the claim here is that it's the actual past, not the agent's current memory appearances or beliefs about the past that make the difference.²²

Some brand of liberated mentalism is, to my mind, the most plausible form of internalism; it rejects EE and SI, after all. However, it doesn't support NDP against reliabilism. It gets the "right" answer concerning demonworlders, but for the wrong reasons.

What makes NDP initially compelling is the intuitive plausibility of the idea that anything that shares my experiences and beliefs is justified in whatever beliefs as I am. Generalizing to the third person, we get constrained mentalism:

(CM): If x and y are experientially and doxastically identical, then they are justificationally identical.

I too feel the intuitive pull of CM, although I find the first-person version significantly more attractive, and as I argued in section 3, the reasons for denying CM outweigh the pull.

Reliabilism faces NDP because reflection on demon scenarios convinces us that CM is true, and reliabilism is incompatible with CM. But the liberated mentalist is in the same boat, since liberated mentalism expressly denies CM and therefore faces NDP objections of its own.

Liberated mentalism claims that mental factors outside the agent's (current) beliefs and experiences affect the justificational status of that agent's beliefs. But this means that I will have an experiential/doxastic counterpart who differs with respect to these other mental factors (learning history, cognitive modules, etc.) and therefore is not justified in beliefs that I am justified in. In general, there will be pairs of introspectively identical agents, who differ with respect to which beliefs are justified for them. This is counterintuitive. But worse, I submit, it is

²² Sarah Wright (2010) holds this view for virtue-theoretic reasons, and Richard Feldman tells me (personal communication) that he is inclined to believe it (see the postscript to "Evidentialism" in Conee and Feldman 2004).

counter to the very intuition that propels the New Evil Demon Problem for reliabilism, the intuition in favor of CM. For here is a possible agent who is exactly like me in all phenomenological and introspectible respects, but isn't justified in any of his beliefs. Thus, the liberated mentalist should be the last person to wield NDP against reliabilism or in favor of her own view.

I'm not sure how to prove that it's really *CM* that serves as the intuitive lynchpin of NDP. One could always try to argue that the plausibility of NDP derives from an intuition in favor of the justificational equivalence of some broader class of cognitive counterparts. Consider a brain supervenience thesis:

(BS): If two agents have (synchronically and diachronically) molecularly identical brains, then those agents are justificationaly identical.

BS does have some intuitive plausibility, and it would support a brain-in-a-vat version of NDP. But the intuitive plausibility of BS seems entirely parasitic on its being a generalized version of CM. Note, in fact, that it is only a generalization of CM (i.e., CM entails BS) on the assumption of semantic and phenomenal internalism. Reject those assumptions and BS retains little or no intuitive plausibility. Reinststate those assumptions, however, and BS has at least as much plausibility as CM, since the former is obviously a safer claim than the latter. In just this way, liberated mentalism may be at least as plausible as CM (since it's a safer claim), even though it derives all of its intuitive plausibility from CM. If this is right, then the truth of liberated mentalism would preclude NDP from serving as evidence of liberated mentalism or anything else. Liberated mentalism would have to get its plausibility from someplace else.

The liberated mentalist probably won't concede all of this, and we're probably at an impasse here. Perhaps it is enough to reiterate that liberated mentalism has its own demons to worry about and thus has no clear advantage over reliabilism as far as NDP goes.

6. Conclusion

I have argued that the New Evil Demon Problem is no embarrassment for in situ reliabilism. There are plenty of details to work out, but the epistemic status of demonworlders is no cause to abandon the central insights of reliabilism.

I have dwelt on the dialectic of the NDP argument more than I would ordinarily want to, because the main thesis here is not so much about the truth of a certain form of reliabilism but about the failure of a certain type of objection to reliabilism. Thus, when certain unobvious, arcane, controversial philosophical positions arose as ways of salvaging the claim that demonworlders are justified in the same beliefs we are, I pointed out that the purpose of NDP was to show that reliabilism is false, not to show that it is at odds with some controversial internalist epistemology. It really does matter that NDP is supposed to work much like Gettier's (1963) famous counterexamples, regarding which we are supposed to have strong intuitive convictions, convictions which are independent of our antecedent epistemological commitments. Any unbiased outsider should be able to see that the cases tell against the traditional analysis of knowledge. But if the only way to hold on to the NDP intuition that our demon world counterparts are justified is to invoke subjectivism, or the epistemological theory of concepts, or the like, the game is already over. Obviously if one of those is true, reliabilism is false, but just as obviously there's no reason to assume that the casualty of this conflict will have to be reliabilism;

the choice (between, e.g., reliabilism and subjectivism) will have to be made on the basis of the overall successes of the respective theories.

This emphasis on the dialectic of NDP is only part of the response to NDP, however. The meat of the response is that reliabilism really does give a satisfactory account of the epistemic status of demonworlders. Their basic, or belief-independent, beliefs are unjustified, and in fact, have no positive epistemic status at all. But this is the intuitively correct verdict after all. When we stop thinking of the demonworlders as *us* in uncooperative environments and take seriously the (constrained) mentalism NDP requires, we see the now unanchored demonworlders as mere cognitive mappings. It is wildly implausible, and inconsistent with NDP's mentalism, to allow that all cognitive mappings generate justification; it is unprincipled and arbitrary to claim that only some of them do (and chauvinistic if these happen to include pretty much just *ours*); the only option left is to deny that any unanchored demonworlders' cognitive mappings yield justification. Because they are in a demon world, and because they have no special connection to any other world, there is nothing to make one cognitive mapping *better* than another, nothing to make one mapping minimally good.

Thus, the reliabilist claims that the demonworlders are unjustified in many of their beliefs, but this is not a matter of biting the bullet, for the claim that our demon world counterparts are not justified in their perceptual beliefs ceases to be counterintuitive.

Nevertheless, I do not want to claim that no demon world beliefs have any positive epistemic status. Nor do I have to. Reliabilism has the resources to ascribe a positive epistemic status to a number of beliefs formed even in environments that are generally not conducive to truth. Some processes will be reliable even in these environments, and others will be

conditionally reliable; these will give us justified and conditionally justified beliefs, respectively. So even though the demonworlder's perceptual beliefs are unjustified (and not even conditionally justified), her memories and inferences might have positive epistemic status nonetheless; the reason they fall short of (categorical) justification is the verdict about perception just embraced.

The response to NDP, therefore, is partly conciliatory (where belief-dependent processes are concerned) and partly defiant (where belief-independent processes are concerned).

NDP remains a problem of detail for reliabilism. I claimed early on that in situ reliabilism makes up a family of theories, and there is a lot of flexibility regarding how narrowly to read the 'situ'. Evil demon considerations can be put to use by helping us decide which understanding of 'in situ'---where the agent is/has been/will be---to endorse. The connection between truth and justification will be complex and indirect---the treatment of belief-independent and belief-dependent processes already ensures this. But better complex than virtually nonexistent, as the connection is on internalist theories and some competing versions of reliabilism. The notion of anchoring has been central, and this will also need to be spelled out. It is hoped that these two problems can be solved concurrently; the analogy to skills seems promising here, but the details will have to wait. In the meantime, reliabilists can press the defiant angle just mentioned: it is, on further reflection, intuitively plausible that unanchored demonworlders are *not* justified in their belief-independent beliefs. In situ reliabilism, and perhaps certain other forms of externalism, can account for this. The onus is now on the opponent of reliabilism to accommodate this result within an internalist framework.

Acknowledgements:

A version of this paper was presented at the Feldmania! conference at the University of Texas, San Antonio, in February, 2011; thanks to the audience there, especially Sandy Goldberg, Richard Feldman, Sarah Wright, Matthew McGrath, Jeremy Fantl, Jonathan Kvanvig, and Clayton Littlejohn, for helpful comments. Special thanks to Tom Senor for comments on written drafts of this material.

References:

- BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge: Harvard UP.
- (2010). *Epistemology: Classic problems and contemporary responses*. 2nd ed. Lanham, MD: Rowman and Littlefield.
- Cohen, S. (1984). Justification and truth. *Philosophical Studies* 46, 279-95.
- Comesaña, J. (2002). The demon and the diagonal. *Philosophical Studies* 110, 249–66.
- (2010). Evidentialist reliabilism. *Nous* 44, 571--600.
- Conee, E. & R. Feldman (2004). *Evidentialism*. Oxford: Oxford University Press.
- DeRose, K. (2000). Ought we to follow our evidence? *Philosophy and Phenomenological Research* 60, 697-706.
- Dretske, F. (1996). Phenomenal externalism or If meanings ain't in the head, where are qualia? *Philosophical Issues* 7, 143-158.
- Engel, M. (1992). Personal and doxastic justification in epistemology. *Philosophical Studies* 67, 133-150.
- Feldman, R. (2003). *Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Feldman, R. and E. Conee (1985). Evidentialism. *Philosophical Studies* 48, 15-34.
- Foley, R. (1987). *The theory of epistemic rationality*. Cambridge: Harvard University Press.
- Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis* 23, 121-3.
- Goldman, A. I. (1979). What is justified belief? In George Pappas, ed. *Justification and knowledge*. Dordrecht: Reidel.

- (1986). *Epistemology and cognition*. Cambridge: Harvard UP.
- (1988). Strong and weak justification. In *Philosophical Perspectives*, 2, Epistemology. J. E. Tomberlin, ed.
- (1992). Epistemic folkways and scientific epistemology. In *Liaisons*. Cambridge, MA: MIT Press.
- Haack, S. (1993). *Evidence and inquiry: Toward reconstruction in epistemology*. Oxford: Blackwell.
- Henderson, D. and T. Horgan (2001). Practicing safe epistemology. *Philosophical Studies* 102, 227–258.
- (2006). Transglobal reliabilism. *Croatian Journal of Philosophy* 6, 171–195.
- Huemer, M. (2001). *Skepticism and the veil of perception*. Lanham, MD: Rowman and Littlefield.
- Lackey, J. (2005). Memory as an epistemically generative source. *Philosophy and Phenomenological Research* 70, 636–658.
- (2007). Why memory really is an epistemically generative source: A reply to Senor. *Philosophy and Phenomenological Research* 74, 209–219.
- Lehrer, K. and S. Cohen (1983). Justification, truth, and coherence. *Synthese* 55, 191–207.
- Lyons, J. C. (2005). Perceptual belief and nonexperiential looks. *Philosophical Perspectives* 19, 237–56.
- (2009a). *Perception and basic beliefs: Zombies, modules, and the problem of the external world*. New York: Oxford University Press.
- (2009b). Perception and virtue reliabilism. *Acta Analytica* 24, 249–61.
- (2010). Reply to critics. *Philosophical Studies*.
- (forthcoming). "Circularity, Reliability, and the Cognitive Penetrability of Perception". *Philosophical Issues*.
- Markie, P. J. (2004). Nondoxastic perceptual evidence. *Philosophy and Phenomenological Research* 68, 530–553.

- Pollock, J. (1974). *Knowledge and Justification*. Princeton: Princeton University Press.
- (1986). *Contemporary theories of knowledge*. Savage, MD: Rowman & Littlefield.
- (1995) *Cognitive carpentry*. Cambridge, MA: MIT Press.
- Pollock, J. L., & J. L. Cruz (1999). *Contemporary theories of knowledge*, 2nd ed. Lanham, MD: Rowman and Littlefield.
- Pryor, J. (2000). The skeptic and the dogmatist. *Nous* 34, 517-49.
- Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.
- Quinton, A. (1966). The foundations of knowledge. In B. Williams & A. Montefiore, eds. *British analytical philosophy*. London: Routledge & Kegan Paul.
- Reid, T. ([1785] 1967). *Essays on the intellectual powers of man*. In *Philosophical works*. W. Hamilton, ed. Hildesheim: Georg Olms Verlagsbuchhandlung.
- Senor, T. D. (2007). Preserving preservationism: A reply to Lackey. *Philosophy and Phenomenological Research* 74, 199-208.
- Sosa, E. (1993). Proper functionalism and virtue epistemology. *Nous* 27, 51-65.
- (2007). *A virtue epistemology*. Oxford: Oxford University Press.
- Tye, M (forthcoming). Phenomenal Externalism, Lolita, and the Planet Xenon.
- Wright, S. (2010). Internalist Virtues and Knowledge. *Acta Analytica* 25, 119-132