New Lessons from Old Demons: The Case for Reliabilism

Thomas Grundmann

Our basic view of the world is well-supported. We do not simply happen to have this view but are also equipped with what seem to us very good reasons for endorsing it. But it still seems possible that the real world is in fact very different from the way it epistemically appears to us, even if the way it appears to us is informed by our best epistemic sources (including the sciences). One radical way of considering the world as being very different from the way it appears to us is to think of our cognitive perspectives as being manipulated by a powerful being, traditionally called "the evil demon", who intends to deceive us about almost everything in the world. Nothing much depends on what exactly we think about the nature, motivation and mode of operation of this demon, e.g., whether we conceive of it as an evil god (as Descartes suggested), as a mad scientist (as in Putnam's seminal brain-in-a-vat case), or as a giant machine (as in the movie *The Matrix*). If some kind of evil demon hypothesis were true of the actual world, our well-established view of the world would be massively mistaken.

What is the epistemic significance of the fact that the evil demon hypothesis might be true of our world, although we currently don't have any evidence that suggests that this possibility is actualized?[1] There are two rather different intuitive responses that are usually addressed separately. On the one hand, many people share what I will call the *Old Evil Demon* (OED) intuition, according to which one must be justified in believing that the evil

---

[1] Pritchard 2012, 126–27, argues that we cannot in principle have empirical evidence that suggests that we are victims of skeptical deception: "Because it is in the very nature of radical skeptical hypotheses that they call one's empirical beliefs into question en masse, it is therefore inevitable that they will call into question whatever empirical basis one takes oneself to have for supposing such a hypothesis to be true." (Ibid.: 127) But this view is implausible. Suppose, e.g., that a voiceover tells us that she is persistently manipulating our experience. She then correctly predicts all kinds of surprising events in our future stream of experience. The best explanation of this pattern of experience would be that what the voice says is true. Hence, contrary to what Pritchard assumes, we do not need objective empirical evidence to support the claim that we are deceived by some powerful being. We can rely on our introspective knowledge about our experience, which is not called into question by the skeptical hypotheses under consideration.

demon hypothesis is false in order to have any justified beliefs about the external world. Descartes maintained a similar view with respect to knowledge. In order to know that p, Descartes claimed, one has to rule out all propositions that are incompatible with the truth of p, including the hypothesis of radical demon deception. Currently, BonJour takes such a requirement as obviously applying to epistemic justification: "to admit that no response to the sceptic is possible […] destroys the claim of […] epistemic justification […] in the first place." (BonJour 1985: 14) Although not everyone will share BonJour's intuition, it seems to be accepted by a significant number of epistemologists and non-philosophers. On the other hand, there is the currently very prominent *New Evil Demon* (NED) intuition, according to which we can have justified beliefs about the external world even if the evil demon hypothesis of massive deception is *actually true*. Many people feel quite attracted to the fallibilist idea that even a false belief can be justified, and the NED intuition looks like a plausible generalization of justificatory fallibilism. Here are examples of two current proponents of the NED intuition. Cohen claims: "It strikes me as clearly false to deny that under these circumstances [i.e., if the mental perspective in the demon world looks fully normal; TG] our beliefs could be justified." (Cohen 1984: 281) Similarly, Foley (1985:190) claims: "Even if […] our world is w [i.e., a demon world; TG], it still can be epistemically rational for us to believe many of the propositions we do […]."

The exact relationship between the OED intuition and the NED intuition is only rarely addressed in the literature. Are these intuitions compatible? And if they are, what meta–epistemological view of justification do they suggest? To the extent that the relation between the two intuitions is explicitly addressed, the prevailing view seems to comprise the following two claims (compare Pryor 2000, BonJour 2010):

(1) The NED intuition suggests *mentalism*, i.e., the view that justificatory properties are fully determined by facts about non-factive mental states of the cognizer.[2] If mentalism is true, justificatory properties do not depend on any facts external to one's mental perspective. Given that a belief-producing mechanism is reliable only if it produces mostly true beliefs in the actual and close possible worlds, mentalism implies that beliefs need not be reliably produced in order to be justified. Hence the NED intuition seems to suggest that reliability is not necessary for epistemic justification.

(2) The OED intuition is compatible with mentalism, i.e., even if justification does not require reliability, there is an explanation why having justified beliefs about the world requires one to be justified in believing that the evil demon hypothesis is false.

In this paper, I will argue for the contrary of both (1) and (2). In the next section, I will argue that at least some variants of the OED intuition strongly suggest that justification implies reliability after all. In the following section, I will argue that, interestingly, this is compatible with the NED intuition, or, at least, with something very near to it. I will conclude by taking a somewhat broader perspective on the results of this paper.

1. Why the OED intuition suggests that justification requires reliability

Isn't there an obvious explanation of why one must be justified in believing that the evil demon hypothesis is false in order to have justified beliefs about the external world — an explanation that is completely independent of any reliability requirement and that is fully compatible with epistemic internalism? The explanation runs like this: Considering the evil demon hypothesis constitutes a defeater to one's *prima facie* justified beliefs about the external world. Having the impression of massive deception "from the inside" defeats our *prima facie* justification even if justification does not require objective reliability. This

---

[2] Until more recently, all proponents understood mentalism in this way, e.g., Feldman & Conee 2004. There are now some dissident views that include factive mental states such as perception (e.g., Pritchard 2012) or knowledge (Williamson 2000) in the mentalist basis.

defeater will remove the justification of one's beliefs about the world unless it is itself defeated by evidence against the demon hypothesis.[3] However, considering the mere possibility of a scenario of massive deception does not suffice to constitute a defeater. This is because regarding a proposition as possible neither amounts to believing that it is true (as doxastic defeaters require) nor does it provide us with any evidence for its truth (as normative defeaters require).[4]

On the other hand, one might think that there is a very short argument that establishes the reliability requirement for justification based on the OED intuition: Some skeptical hypotheses (SH), e.g., the dreaming hypothesis, are incompatible with the reliability of our beliefs about the world; if we need good reasons for ruling out such hypotheses in order to be justified in our beliefs about the world, as the OED intuition requires, then (so the argument goes) there cannot be justified beliefs about the world unless we are in fact reliably connected to the world. However, this argument suffers from a severe defect. It presupposes that "having good reasons for ruling out SH" means the same as "having *infallible reasons* for believing that SH is false". A more charitable reading of the OED intuition, however, only requires that one has sufficiently good reasons for believing that SH is false. As long as one is fallibilist about justification, one can be justified in believing a proposition that is in fact false. The OED intuition is thus compatible with the idea that even an inhabitant of a demon world may justifiedly (but not infallibly) believe that the demon-world scenario does not apply to the actual world. Therefore, the OED intuition does not seem to require that justification involves reliability.

Further, one might doubt that the OED intuition can be explained unless justification requires reliability.[5] Specifically, why should one try to justify the belief that demonic

[3] This internalist explanation of the OED intuition was suggested to me by David Enoch.
[4] For a comprehensive discussion of epistemic defeaters, see Grundmann 2011.
[5] In what follows, it is important to keep in mind that the order of justification differs from the order of explanation. The justification starts with immediate intuitions about demon worlds. From here we infer to explanatory principles from which what is intuited can be deductively derived. The explanations I offer below

hypotheses are false if one's intention is not to justify the belief that one is reliably connected to the world as is required by epistemic justification? There is, however, an answer to the question of why one should try to justify the belief that demonic hypotheses are false that is completely independent of a reliability requirement for justification. Consider a plausible version of the closure principle for justification: *in order to be justified in her ordinary belief that p, an epistemic agent must be justified in believing that all known alternatives to p are false.*[6] Whenever we consider evil demon hypotheses we quickly realize, i.e., come to know, that some of them are incompatible alternatives to the propositions we ordinarily assume ourselves to be justified in believing. Consider my ordinary claim that I am justified in believing that *I have two hands*. Having two hands is obviously incompatible with either living in a demonic void or being a handless brain in a vat. Hence, in order to be justified in believing that I have two hands, the closure principle requires me to be justified in believing that I am neither living in a demonic void nor a handless brain in a vat. Thus, there is a good and straightforward explanation of our OED intuition that does not imply that justification requires reliability.

On my view, there is a better argument for claiming that the OED intuition calls for an externalist reliability condition on epistemic justification. But the argument is more complex than the ones just considered. To explicate this argument, we need to distinguish two different kinds of skeptical hypotheses.[7] Skeptical hypotheses of the first type (SH1) are hypotheses that are incompatible with the *truth* of everything we ordinarily believe about the world. The best version of such a *hypothesis of global deception* seems to be Descartes' hypothesis of a demonic void without any material external world. This hypothesis of global deception must be distinguished from skeptical hypotheses of the second type (SH2), which describe

---

are not meant to be inferential justifications of what is intuited. The relation between justification and explanation is similar to that in hypothetico-deductive methods.

[6] See, e.g., Wright 1985.

[7] For this distinction, see Grundmann 2003: 71–72.

scenarios in which *the sources of our beliefs about the world are epistemically inadequate*. While dreaming or hallucinating one seems to lack good perceptual reasons for what one dreams or hallucinates, even if what one dreams or hallucinates is true—which can surely be the case.[8] Consider dreaming that one is sleeping in one's bed or dreaming about one's actual name or profession. Therefore, SH2—which I will generically call the "dreaming hypothesis"—conflicts with one's ordinary belief being justified rather than with it being true.

Given this distinction between skeptical hypotheses that are incompatible with the *truth* of our ordinary beliefs and skeptical hypotheses that are incompatible with the *justification* of what we ordinarily believe, let us reconsider the original OED intuition, which claims that in order to have justified beliefs about the world, we must be justified in believing that the relevant skeptical hypothesis is false. When one takes into account that there are two different types of skeptical hypotheses (SH1 and SH2), it seems quite natural to distinguish the following two different readings of the OED intuition:

OED*: In order to have justified beliefs about the world, we must be justified in believing that the *skeptical hypothesis of global deception* about the world is false.

OED**: In order to acquire justified beliefs about the world on the basis of sense experience, we must be justified in believing that the *dreaming hypothesis* is false.

It is easy to show that OED* can be explained without relying on the premise that justification involves reliability. We can give the required explanation by way of the following deductive argument:

(1)   For all p, in order to be justified in believing that p we must be justified in believing that all things we know to be alternatives to p are false.

(2)   We know that the skeptical hypothesis of global deception about the world (SH1) is an alternative to whatever we believe about the world.

Therefore,

---

[8] See, e.g., Moore 1959: 245–46; Stroud 1984: 16–17; Wright 1986: 432–33.

OED*   In order to have justified beliefs about the world, we must be justified in believing

that the *skeptical hypothesis of global deception* about the world is false.

This argument is clearly valid. Premise (2) seems obvious, since SH1 is defined as being incompatible with the truth of everything we believe about the world. Premise (1) is a version of the *closure principle* for justification. So the argument is sound. But then no premise about the reliability of justification is needed to explain OED*.

Consider, in contrast, OED**. This intuition cannot be explained in a way analogous to OED*. Why not? Because, as I said before, dreams may and sometimes do correspond to reality. Hence, the closure principle does not require that one is justified in believing that the dreaming hypothesis is false. So, we have to look for some other explanation of OED**. In what follows I will offer what I take to be the best explanation. I won't be able to rule out all possible alternatives. But I will argue that at least the most salient alternatives do not work. Here is one deductive argument that has OED** as its conclusion and hence offers an explanation, E1, of OED**:

(1)   We acquire justified beliefs about the world on the basis of sense experience only if we are justified in believing that we acquire justified beliefs about the world on the basis of sense experience.

(2)   We know that one cannot acquire justified beliefs about the world on the basis of sense experience when one is dreaming.

(3)   If we are justified in believing that we acquire justified beliefs about the world on the basis of sense experience and if we know that one cannot acquire justified beliefs about the world on the basis of sense experience when one is dreaming, then we are justified in believing that we are not dreaming that p.

Therefore,

OED** In order to acquire justified beliefs about the world on the basis of sense experience, we must be justified in believing that we are not dreaming.[9]

What is the rationale behind the premises of this valid argument? Premise (1) expresses a kind of JJ–principle for propositional justification. In order to acquire justifiers (of a certain kind), we need to have good reasons that we acquire those kinds of justifiers. I will discuss the plausibility of this premise below. Premise (3) expresses a special instance of the previously introduced closure principle for justification. What about premise (2)? If one knows that one cannot acquire justified beliefs on the basis of sense experience when one is dreaming, then one cannot acquire justified beliefs on the basis of sense experience when one is dreaming. This is so because knowledge is factive. But why should it be impossible to acquire the justified belief that p on the basis of dreaming that p? One plausible answer is that (i) one's sense experience is unreliable as long as one is dreaming and (ii) one can acquire new justification only on the basis of reliable sources. The upshot of the above argument is that this explanation of OED** presumes that justification requires reliability.

The OED** intuition is sometimes used in skeptical arguments against the empirical justification of our beliefs about the world.[10] Here is a schema for such arguments, driven by the OED** intuition:

(P1)   In order to acquire justified beliefs about the world on the basis of sense experience, we must be justified in believing that the dreaming hypothesis is false.

(P2)   We cannot be justified in believing that the dreaming hypothesis is false.

Therefore,

---

[9] More formally, the argument runs like this:
(1) $J_e p \rightarrow JJ_e p$
(2) $K(J_e p \rightarrow \neg Dp)$
(3) $(JJ_e p \ \& \ K(J_e p \rightarrow \neg Dp)) \rightarrow J(\neg Dp)$
Therefore,
OED** $J_e p \rightarrow J(\neg Dp)$
The capital J stands for *justification*, with the index $_e$ it stands for *empirical justification*, K stands for *knowledge*, and D for *dreaming*.
[10] See, e.g., Wright 1991; Grundmann 2003: 216–26.

(C)   We cannot acquire justified beliefs on the basis of sense experience.

Premise (P1) is equivalent to the OED** principle. If we accept this principle, the force of the skeptical argument depends entirely on premise (P2). Opinions differ on this premise. Some argue that one surely can have good empirical reasons against the dreaming hypothesis.[11] Doing so commits one to a kind of bootstrapping or epistemically circular justification. If one does not accept any of these justificatory strategies, one might still claim that one can have good *non-empirical* reasons against the dreaming hypothesis: either a priori reasons or some kind of default justification.[12] So, even if OED** is a crucial step in the skeptical argument against empirically justified beliefs about the world, it is not by itself sufficient to generate justificatory skepticism.

So far, I have argued that there is not a unified OED intuition, but rather two different intuitions that claim that being justified in believing that the evil demon hypothesis is false is necessary for being justified in believing ordinary propositions on the basis of sense experience. One of these intuitions, OED*, can be fully explained on the basis of closure. But to make sense of the other intuition, OED**, we apparently have to assume that epistemic justification requires reliability. Proponents of mentalism might respond that it is not clear that we all share the OED** intuition. But even if we do have this intuition, it might be explained away by claiming that the primary intuition about the old evil demon is OED*, and that we somehow do not distinguish clearly enough between this case and the dreaming hypothesis because dreams often falsely represent the environment. Hence we confuse OED* with OED**. If this were true, we would not need any additional explanation of OED**, for it would then be shown to be untrustworthy.[13] But this strategy would seem to be a desperate

---

[11] For example, Austin 1962: 49 claims that dream experience is qualitatively different from waking experience. For a more timely defense that relies on experimental dream research, compare also Grundmann 2002. Similar views were held by Descartes, Hobbes and Locke.

[12] For an a priori argument against skeptical hypotheses, see BonJour 1985, Ch. 8. Wright 2004 employs the idea of default justification against the skeptical challenge.

[13] Something along these lines was suggested to me by Anthony Anderson.

move. Given that many epistemologists still find OED** intuitively appealing, we can safely ignore this move here.

Here is a more substantial objection to the explanation of OED** suggested above. One might argue that there is a much simpler and more straightforward way of explaining OED** than the one I suggested (i.e., E1):[14]

(1)   If we are justified in believing that p on the basis of sense experience and if we know that if p is true and we are experiencing that p, we are justified in believing that we are not dreaming that p, then we are justified in believing that we are not dreaming that p.

(2)   We know that if p is true and we are experiencing that p, we are justified in believing that we are not dreaming that p.

Therefore,

OED** In order to be justified in believing anything on the basis of sense experience, we must be justified in believing that the *dreaming hypothesis* is false.

Let us discuss the premises in reverse order. Premise (2) seems plausible since it is very likely that true experiential representations do not constitute dreams. It is true that dreams can be factive, but only accidentally. Premise (1) expresses an instance of a transitivity principle for justification. More formally: $Jp \ \& \ K(p \rightarrow Jq) \rightarrow Jq$.[15] However, transitivity for justification is in danger of being too permissive. To see the worry more clearly, suppose that the justification for my belief is constituted by the degree of probability of the believed proposition given the evidence available to me. We may stipulate, for the sake of argument, that the probability of a belief must be higher than .6 for the belief to count as justified in the absolute sense. Now suppose that an epistemic agent A believes that p with a probability of .7 and also believes that the conditional probability of q given p equals .7. If she infers from this

---

[14] This objection was raised by Marian David. David Enoch helped me to make the objection as strong as possible.
[15] The capital J stands for *justification*, K stands for *knowledge*.

to q, her conclusion has a probability of .49. This is below the threshold for justification in the absolute sense. So, this case constitutes a counterexample to the general principle of transitivity for justification. Of course, this is not a knockdown argument, since one might spell out justification in non-probabilistic terms.[16]

So far it has been argued that among the OED intuitions there is at least one intuition, namely OED**, whose best explanation involves a reliabilist constraint on epistemic justification. If one looks more carefully at this explanation, it turns out that it depends in part on a highly controversial principle, the JJ–principle (premise 1 in E1). This principle is problematic for at least two reasons. First, if one always needs some kind of meta-justification for any kind of first-order justification, this looks like an over-intellectualized account of justification. Second, the JJ–principle directly leads to a vicious regress.[17] So the best explanation seems to be no good explanation after all. But then we are at the following impasse: there is an epistemic intuition, OED**, that can only be explained by epistemic principles that are, at least in part, dubious. For this reason, I won't rest my further argument on this version of the OED intuition and will bracket OED** for the rest of this paper.

The interesting question now is whether there is any intuition in the vicinity of OED** (i) that does not rely on the highly controversial JJ–principle and (ii) that cannot be explained without assuming that justification requires reliability. On my view there is indeed such an intuition. Consider whether your current perceptual beliefs about the world are justified. When you ask yourself this, you treat it as an open question, i.e., you are looking for reasons to decide the issue that do not presuppose any specific answer. Moreover, when you ask this question, you are looking for justifying reasons in order to decide whether your beliefs have first-order justification. How can you answer this second-order question positively?

---

[16] This reply was suggested to me by Joachim Horvath.
[17] Compare Alston 1989c.

Intuitively, it would seem absurd to argue that you have justified perceptual beliefs just by referring to your *introspective knowledge* that your perceptual beliefs correspond to your sense experiences. You cannot tell whether your beliefs about the world are justified on the basis of nothing but introspection. When you rationally claim that you have justified perceptual beliefs you seem to be committed to the further claim that you are currently not dreaming. So, in order to give any justified affirmative answer to the question of whether you have justified perceptual beliefs, you need to be justified in believing that you are currently not dreaming.[18] This, then, is the intuition we have been looking for:

OED***: If you are justified in believing that you have acquired justified beliefs about the world on the basis of sense experience, you need to be justified in believing that you have not been dreaming.[19]

We can explain this intuition without appealing to any controversial JJ–principle. Here is the explanation:

(1)   You know that you cannot acquire justified beliefs about the world on the basis of sense experience when you are dreaming.

(2)   If you are justified in believing that you have acquired justified beliefs about the world on the basis of sense experience and if you know that you cannot acquire justified beliefs about the world on the basis of sense experience when you are dreaming, then you are justified in believing that you have not been dreaming.

Therefore,

OED*** If you are justified in believing that you have acquired justified beliefs about the world on the basis of sense experience, you need to be justified in believing that you have not been dreaming.[20]

---

[18] Compare Grundmann 2003: 265.

[19] OED*** does not claim that you have to justify that you are not dreaming *before* or *independently* of any justified beliefs about the world. It is, e.g., compatible with the view that you justify that you are not dreaming on the basis of bootstrapping from your justified first-order beliefs about the world.

[20] Here again is a more formal version of the argument:

The OED*** intuition can obviously be explained without the dubious JJ-principle. We just need an instance of justificatory closure under *known* implication (premise 2) and the assumption that we know that we do not acquire perceptual justification merely by dreaming (premise 1).

Is it really true that the best explanation of premise (1) is that justification requires reliability? I want to defend this explanation by way of considering four different explanations of why the acquisition of empirically justified beliefs is incompatible with dreaming that do not depend on the assumption that justification requires reliability. Here is a *first* such explanation: dreaming states are neither perceptual nor experiential but are states of imagining. There are many differences between perception and imagination.[21] In contrast to sense perception, imagination is, e.g., not causally responsive to stimuli in the environment and is under our voluntary control.[22] Hence, dreams cannot provide us with *perceptual* justification for our beliefs about the world. Moreover, imagination lacks the affirmative character of representation, i.e., the property of representing its content *as being actually true*, which is necessary for being a justifying reason.[23] Being a mental state with the relevant (propositional) content is clearly not enough for justifying one's belief. This is the reason why the *desire* that a particular person is around or the *supposition* that she is around cannot justify the belief that she is around. Some philosophers argue that dreams lack the required affirmative character as well. Although, according to these philosophers, dreams have an

---

(1) $K(J_e p \rightarrow \neg Dp)$
(2) $(JJ_e p \ \& \ K(J_e p \rightarrow \neg Dp)) \rightarrow J(\neg Dp)$
Therefore,
OED*** $JJ_e p \rightarrow J(\neg Dp)$
The capital J stands for *justification*, with the index $_e$ it stands for *empirical justification*, K stands for *knowledge*, and D for *dreaming*.

[21] For a comprehensive discussion of these differences, see McGinn 2004: 7–41.

[22] McGinn 2004: 12–17.

[23] McGinn (2004: 21): "Percepts supply (defeasible) reasons to believe; they insist on their own veracity. But images do not invite belief in this way; they do not purport to tell us how the world is. […] The image is not evidence that things are presently thus-and-so in the external world […]. Percepts entitle you to form the corresponding belief, but images do not."

immersive quality, they do not strictly possess an affirmative character. For this reason they cannot provide us with *genuine justification* for our beliefs about the world.[24]

Admittedly, this explanation might succeed in explaining why a belief cannot be empirically justified by a dream with the same content, and it does so without presupposing that justification requires reliability. But we can easily find another intuition in the vicinity of the OED*** intuition such that what has to be ruled out by good reasons is hallucination rather than dreaming. OED***+ reads as follows: *If you are justified in believing that you have acquired justified beliefs about the world on the basis of sense experience, you need to be justified in believing that you have not been hallucinating.* What might explain why dreams cannot provide us with justification for our beliefs about the world in non-reliabilist terms does not serve to also explain why hallucinations cannot provide us with such a justification. The alternative explanation does not apply to OED***+ since, e.g., visual hallucinations are clearly sense-experiential states with an affirmative character.

Here is a *second* putative explanation of why we have to be justified in believing that the dreaming hypothesis is false in order to acquire empirical justification for our beliefs about the world. One might think that the dreaming hypothesis constitutes a certain type of defeater that removes one's *prima facie* empirical justification unless it is defeated itself. According to this view, the *dreaming hypothesis* would constitute an undercutting defeater (which conflicts directly with the efficaciousness of the prima facie justification), whereas the *skeptical hypothesis of global deception* would constitute a rebutting defeater (which conflicts with the truth of the prima facie justified beliefs). The nice thing about defeaters is that what they represent need not be true in order to make them epistemically efficacious. It will do if a defeater is simply *believed* or if there is some *evidence that suggests that the defeater is true*. But if defeaters need not actually be true in order to do their job, then even an epistemic

---

[24] See McGinn 2004: 96–112; Sosa 2007.

internalist who does not believe that justification requires objective reliability can make use of them.

However, considering the *mere possibility* of a scenario that would be in conflict with either the justificatory efficaciousness or the truth of a given belief does not suffice as a defeater.[25] Since considering the dreaming hypothesis does not generate any kind of defeater, it is not necessary to defeat this hypothesis in order to remain justified in one's beliefs about the world.

At this point, one might concede that there is no fully internalist explanation of the OED*** intuition. Somehow we must refer to external requirements for justification in order to explain why acquiring empirical justification and dreaming (or hallucinating) are incompatible. But there might still be an externalist account of this explanandum that does not involve a reliabilist condition.[26]

Something like the following *third* explanation suggests itself: empirical justification originates from factive mental states like perceptions of facts or knowledge.[27] Not every true mental representation automatically qualifies as a factive mental state. For this reason, true dreams or veridical hallucinations do not count as justifiers; something more is needed to constitute genuinely factive mental states. This account seems to motivate the claim that neither dreams nor hallucinations can provide justification, yet without being committed to the stronger view that justification involves reliability.

One might object that an account of factive reasons makes it impossible to account for justified false beliefs. But this is not correct. Even if the initial reasons must involve facts, their cognitive processing might still result in false beliefs if the beliefs do not perfectly correspond to the initial reasons' content. This may happen if what is perceived is falsely described or if one draws a content-extending inference, e.g., an inference to the best

---

[25] This has already been argued above. See pp. ???[3]–???[4].
[26] This objection was suggested to me by Tim Kraft in personal conversation.
[27] For different versions of this account, see Williamson 2000 and Pritchard 2012.

explanation.[28] There is, however, a further, crucial objection. The proponents of factive reasons (such as Pritchard and Williamson) claim that factive mental states are more than just true representations. But what is the extra component that turns true representations into factive mental states? Both Pritchard and Williamson claim that the extra component of factive mental states is an epistemic one. Pritchard says: "Seeing that p, like knowing that p, expresses a rather robust epistemic relation that one bears to p." More specifically, Pritchard claims that "seeing that p could constitute one's epistemic basis for knowing that p",[29] and Williamson maintains that knowledge implies safety.[30] On both accounts, the relevant factive mental state thus invokes an anti-luck condition that involves some kind of reliability constraint. We can now see more clearly that accounts of factive reasons are themselves presumably committed to a reliability requirement and therefore do not establish a genuine alternative to the explanation I have suggested.

Against the reliabilist explanation of the OED*** intuition, one might *finally* claim that the incompatibility of dreaming and hallucinating with acquiring empirical justification does not require any explanation at all. That is, one might argue that dreaming and hallucinating are simply paradigmatic cases of states that prevent epistemic agents from acquiring new pieces of evidence. An answer to the question *why* dreaming and hallucinating are in conflict with acquiring new justifiers would thus not be required. And if no explanation for OED** is required, then no *reliabilist* explanation is required either.[31]

However, it seems hard to believe that it is a brute and unexplainable fact that dreaming and hallucinating are incompatible with acquiring new justification. It is very natural to assume that both cases must have something in common by virtue of which they

---

[28] The latter alternative is endorsed by Williamson 2000.
[29] See Pritchard 2012, p. 21.
[30] See Williamson 2000.
[31] Something along these lines was independently suggested to me by Jochen Briesen, Jim Pryor and Timothy Williamson in personal conversation.

cannot provide empirical justification. To insist that no explanation is required or even available here seems like a fairly desperate move.

Finally, we have reached the desired result: There is an OED intuition, namely OED***, whose presumably best explanation requires us to assume that justification implies reliability. I must concede that I have not provided a conclusive defense of this claim. Rather, I have argued that all alternative explanations that I have considered here do not succeed. I take it that the burden of proof is now on my opponent's side.

2. Why the NED intuition does not suggest that justification is independent of reliability

If the results of the previous section are basically correct, a plausible version of the Old Evil Demon intuition suggests that justification requires some kind of reliability. However, the OED intuitions are not the only kind of intuitions we have with respect to demonic worlds. There is also the New Evil Demon intuition, i.e., the intuition that victims of demonic deception can still have justified beliefs, which suggests, at least according to the received view, that justification is *independent* of reliability.[32] Hence, we seem to have conflicting intuitions with respect to demon worlds. In the face of this worry, I will take a closer look at the NED intuition in this section.

To make some progress here, we should explicate the content of NED in more detail. According to a common view, the core of this intuition is that justificatory facts supervene on, or are determined by, facts about the non-factive mental perspective of the epistemic agent. Hence, we get as a first approximation:

NED*: The external world beliefs of our mental duplicate in the demon world would have the same justificatory status as our beliefs in a normal world.[33]

---

[32] See, e.g., Foley 1985.
[33] For a similar view, see Pritchard 2012: 38.

Spelled out in this way, NED does not entail that the beliefs of our mental duplicate in the demonic world are actually justified. This would clearly beg the question against the skeptic. It might turn out that *we*, although living in a normal world, don't possess justified beliefs about the world because we do not satisfy certain epistemic requirements. If this were true, our duplicate would also lack justified beliefs according to NED*. Now, our beliefs in the actual world are *ex hypothesi* formed by reliable processes such as perception that function unreliably in the demonic world. Since, according to NED*, there is no difference in justificatory properties between us and the demon's victim, being reliably produced may not be sufficient for justification. Hence, NED* may challenge simple versions of reliabilism that claim that a belief is justified *if* and only if it is based on a reliable process. But NED* neither shows that reliability is not necessary for justification nor does it establish mentalism as the correct view.

There are other versions of NED intuitions that might do better in calling into question the claim that reliability is necessary for justification. Consider this one:

NED**: The victim of demonic deception *can* be epistemically justified in her beliefs about the world.

For the sake of the argument, I will grant that NED** expresses an intuition about genuine *epistemic justification* rather than an intuition about having blameless, internally rational beliefs or about being the beliefs of a virtuous epistemic agent.[34] BonJour (2010: 228–29) clearly commits himself to something like NED**:

> The evil genius carefully controls their [the victims'] sensory and introspective experience, producing in them just the experiences they would have if they inhabited a particular material world, perhaps one exactly like our own, containing various specific sorts of objects and processes that interact and influence each other in a lawful way. The people in this position are, we may suppose, careful and thorough investigators. They accumulate large quantities of sensory evidence, formulate hypotheses and theories, *subject their beliefs to philosophical arguments [...] for the likely truth of their resulting beliefs*. Are the beliefs about their apparent world that the

---

[34] For weaker interpretations of the NED intuition, see Bach 1985, Engel 1992 and Weatherson 2008.

people in such a Cartesian demon world arrive at in these ways justified? […] From an intuitive standpoint, it seems hard […] to deny that they are.[35]

BonJour not only claims that mental duplicates have the same justificatory status (as NED* claims), but also maintains that the victim of a demonic deception will have justified beliefs about the world if she fulfills all her epistemic obligations and possesses a philosophical meta-justification for her beliefs about the world as NED** requires.

In contrast to NED*, NED** seems to provide us with all that is needed to argue against the reliability condition of epistemic justification. Here is the corresponding *argument against reliability* in its standard form:

(P1)   It is possible that an inhabitant of a demon world has epistemically justified beliefs about the world.

(P2)   Necessarily, *all* belief-forming processes are unreliable in demon worlds.

Therefore,

(C)   Epistemic justification does not depend on reliable belief-forming processes.

At first glance, this argument looks quite compelling. It is clearly valid, (P1) is one way of expressing the NED** intuition, and (P2) seems plausible given that the inhabitant of the demon world is a victim of an all-encompassing demonic deception.

In spite of its initial plausibility, however, (P2) is not correct. At least this is what I want to now argue. In order to see which processes will remain reliable in a demon world, one has to fill in more details about what goes on in such a world. According to the standard stipulation of the case, my demonic counterpart shares his non-factively individuated mental perspective with me. What differs between us are the contingent facts in the external world. All contingently existing external facts are completely removed in the demon world. My counterpart is living in a void. If this is an appropriate characterization of the demon world, then *perception* and *memory* of external states of affairs will function unreliably in the demon

---

[35] My italics. See also Cohen 1984: 28:

world. But there are other cognitive faculties whose reliable functioning would not be at all impaired by the demonic deception.[36] Consider, first, *introspection*. Since (i) *ex hypothesi* I share my mental perspective with my demon-world counterpart and (ii) the actual reliability of introspection is fully determined by my mental perspective, introspection must be reliable in the demon world as well. Secondly, consider *rational intuition*. Again, we start with the stipulation that rational intuition is reliable in the actual world. My demon-world counterpart has the same rational intuitions as I do, because this is implied by the sameness of our mental perspectives. Let us further assume that rational intuitions are only about *metaphysical* modalities. They represent what is *metaphysically* possible and what is *metaphysically* necessary. Metaphysical modalities are stable across all possible worlds. Hence, what is metaphysically possible in the actual world is metaphysically possible in any world, and what is metaphysically necessary in the actual world is metaphysically necessary in any world. This feature of metaphysical modalities is represented by unrestricted accessibility in modal logic. Together, these facts entail that rational intuitions are reliable in demon worlds as well. Even the evil demon cannot change the space of possible worlds, which are the truthmakers of my rational intuitions. Moreover, the rational intuitions in the demon world *ex hypothesi* represent the same modalities as in the actual world. But then, rational intuition must be reliable in the demon world if it is reliable in the actual world. Finally, consider *inference*. Deductive inferences are truth-preserving in all possible worlds. No matter what the world is like, a deductive inference leads one to a true conclusion if one's premises are true. Given that an inferential method is conditionally reliable if and only if it maps true inputs to mostly true outputs, deductive inferences are conditionally reliable even in the demon world. But what about *non-deductive* inferences? Not every non-deductively inferential method is conditionally reliable in demon worlds. Take, e.g., inductive generalizations. If the demon world lacks the requisite uniformity of nature, inductive generalizations are not reliable in

---

[36] See Grundmann 2003: 261–263; Lyons 2013.

demon worlds. However, there are certain kinds of non-deductive inferences that are reliable even in demon worlds. Consider, e.g., the following inference pattern (IP):

(P1)   I represent that p.

(P2)   Representations are reliable.

Therefore,

(C)   p.

Obviously, the conjunction of (P1) and (P2) does not deductively entail the truth of p. But in any world in which (P1) and (P2) are true, the truth of p is (objectively) probable. This is true even for the demon world. Here we have a case of a non-deductive inference that is conditionally reliable even in the demon world, although its premise (P2) is, of course, clearly false in the demon world. As we will see below, this is exactly the type of inference that the demon victim can use to justify her beliefs about the external world.

Consider the victim of demonic deception. All her beliefs about the world are *ex hypothesi* false. Is she still equipped with reliable processes that result in beliefs about the world? Although this seems to be impossible at first glance—remember that the victim's beliefs about the world are all false—there is indeed a Cartesian route available to the victim to form beliefs about the world in a reliable manner. Note that, from a reliabilist point of view, it is sufficient for a belief to be inferentially justified if it based on conditionally reliable inferences from reliably formed beliefs. But then our protagonist could use the following inferential method to arrive at beliefs about the world in a manner that is acceptable to the reliabilist:[37]

(P1)   I experience as if p.

(P2)   Experience is reliable in every possible world.

Therefore,

---

[37] For this argument, see Lyons 2013:7–8. Surprisingly, Lyons' own attempt to square reliabilism with the NED intuition is not based on this argument.

(C)   p.

Premise (P1) is based on introspection, which functions reliably even in the demon world, as I argued above. Premise (P2) is based on rational intuition, which also functions reliably in the demon world. Keep in mind that (P2) is actually false if we suppose that the cognizer is in fact living in a demon world, for experience does not function reliably in the demon world.[38] Therefore, there is at least one world in which experience does not function reliably, and this contradicts the claim that (P2) makes, namely, that experience is reliable in *every* possible world. But even if (P2) is false, it could still be based on a reliable process since reliability does not entail infallibility. You might wonder how (P2) can be justified on the basis of rational intuition. I don't want to defend a particular line of argument here, nor do I believe that there is any successful line of argument available. Since I grant that demon worlds are possible, there cannot be any *sound* argument supporting (P2).[39] But I do think that there are several rationalist attempts to give an a priori defense of something like (P2), e.g., Descartes' proof of a benevolent God that supports the reliability of sense perception, or rationalist arguments from semantic externalism that support a minimal reliability of perceptual beliefs.[40] Finally, the inference from (P1) and (P2) to (C) is non-deductive. Since reliability does not guarantee truth, it is possible that a necessarily reliable process results in a false belief. Nevertheless, the non-deductive inference at hand guarantees that the conclusion is

---

[38] In fact, (P2) is necessarily false, i.e., false in all worlds, if the actual world is a demon world. Thanks to David Enoch for reminding me of this fact.

[39] As long as we stipulate that the inhabitant of the demon world possesses a mental perspective that exactly duplicates my own, this generates a devastating problem for my argument. For I have good reasons to doubt the soundness of the rationalist meta-justification. But if I have these reasons, my demonic counterpart, insofar as he duplicates my mental perspective, possesses them as well. Now, these reasons constitute defeaters for the rationalist meta-justification, so there is no mental duplicate of me in any demon world that has *ultima facie* justified beliefs about the world. But then my general argument seems to fail.
Notice, however, that it is not part of NED** that victims of demonic deception share my mental perspective. NED** only claims that there can be inhabitants of demon worlds that have justified beliefs about the world. In order to defend the view that these agents have reliable processes at their disposal, one must concede that I and my counterpart share our mental perspectives at least in part. Otherwise, the rational intuitions of my demonic counterpart could not be as reliable as my own. But the correspondence of perspectives can be restricted in order to avoid having my demonic counterpart inherit my defeaters for the rationalist meta-justification. Thanks to Joachim Horvath for making me aware of this complication.

[40] The locus classicus for this kind of semantic externalist argument is Davidson 1983. In Grundmann & Misselhorn 2003, we defend a rationalist understanding of semantic externalist arguments against external world scepticism.

objectively probable in every world in which the premises are true. This will certainly satisfy reliabilist standards. Thus, we come to the surprising conclusion that even in a world of demonic deception in which all beliefs about the external world are false, the epistemic agent could justify her beliefs about the world in a way that is acceptable to reliabilist standards. Hence, NED** does not conflict with reliabilism after all.

So far, this is good news for the reliabilist. However, there is a further intuition about new evil demons that does conflict with reliabilism. Consider the following case:

NED***: The epistemic agents in the demon world base their beliefs about the world directly on sense experience, as we typically do. They also have a priori arguments that support the reliability of sense experience at their disposal. But in contrast to the case considered above, these arguments do not play a role in the formation of their beliefs about the world. Still, it is hard to deny that the agents arrive at justified beliefs in that world.

If this is what you intuit about the case, then according to your intuition, the relevant justification does not require reliability.

I agree that the reliabilist cannot fully explain NED***. But she can do something close to it.[41] She can explain how the victim of demonic deception can be justified in believing that her perceptual beliefs are justified—even if her justified second-order belief is false. Hence, the reliabilist can attribute meta-justification, but not first-order justification, to some inhabitants of demon worlds. This reliable meta-justification would be based on reliable rational intuition and conditionally reliable inference, as in the previous case. Now, what is established by NED***? One might want to say that reliabilism is refuted if one takes one's intuitions at face value. Alternatively, one might claim that reliabilism can explain almost all of our intuitions about the demon world. With respect to the unexplained rest, it seems rather unclear whether we really have the intuition that the victim's beliefs are first-order justified or

---

[41] For this strategy, see also Grundmann 2003: 267–69.

whether what we really intuit is that the inhabitant of the demon world is justified in believing that she has justified first-order beliefs.

3. Conclusion

What role intuitions should play in analyzing epistemic categories is highly controversial in current meta-epistemology. Some people believe that surveys of intuitive responses to hypothetical cases suggest that intuition is in general too problematic to qualify as a trustworthy source of analysis.[42] Others think that analysis of our epistemic *concepts* is not a suitable method for discovering the metaphysical structure of epistemic categories.[43] I don't want to enter into this debate here. I will just note that I believe that intuitions about hypothetical cases are an important means to improve our understanding of, e.g., the true nature of epistemic justification. For this reason, I take epistemic intuitions very seriously.[44]

I started out with the observation that we have different intuitions about justification with respect to different skeptical hypotheses. According to the New Evil Demon intuition, which has dominated the recent meta-epistemological debate, one can have justified beliefs about the world even if the skeptical hypothesis is true, i.e., even if one is living in a demon world. According to the Old Evil Demon intuition, which is currently widely ignored within the meta-epistemological debate, one cannot possess justified beliefs about the world unless one is able to rule out relevant skeptical hypotheses. There is a strong tendency in current epistemology to regard the NED intuition as evidence for the internalist view that justification is independent of reliability and supervenes on the epistemic subject's non-factive mental perspective. At first glance, this view seems to be compatible with the OED intuition.

---

[42] A paradigm of this experimental philosophy critique of epistemic intuitions is Weinberg, Nichols and Stich 2001.

[43] For this naturalistic challenge to conceptual analysis, see Kornblith 2002.

[44] For a defense of the use of intuitions as evidence, see Grundmann 2007 and 2010.

On closer inspection, we have found versions of the OED intuition, especially those requiring that we are justified in believing that the *dreaming hypothesis* is false, that cannot be explained on the basis of mentalism about justification but call instead for a robust connection between justification and reliability. This fits nicely with the observation that the NED intuition does not provide a compelling argument for mentalism but is in fact compatible with the view that justification requires reliability. Such a view may not fully explain all variants of NED intuitions; for example, it conflicts with NED***. But reliabilism is at least able to explain why inhabitants of demon worlds can be justified in believing that they have justified first-order beliefs, even if that second-order belief is not true. A general, systematic investigation of our intuitions about the relation between justification and skeptical hypotheses suggests that, all things considered, reliability is a necessary condition for justification after all.

References

Alston, William 1989: "Internalism and Externalism in Epistemology". In: William Alston, *Epistemic Justification: Essays in the Theory of Knowledge*, Ithaca/London: Cornell University Press, pp. 185–226.

Austin, John 1962: *Sense and Sensibilia*, Oxford: Oxford University Press.

Bach, Kent 1985: "A Rationale for Reliabilism". In: *Monist* 68, pp. 246–63.

BonJour, Laurence 1985: *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

Cohen, Stewart 1984: "Justification and Truth". In: *Philosophical Studies* 46, pp. 279–95.

Davidson, Donald 1983: "A Coherence Theory of Truth and Knowledge". In: Dieter Henrich (ed.), *Kant oder Hegel?*, Stuttgart: Klett–Cotta, pp. 423–38.

Engel, Mylan 1992: "Personal and Doxastic Justification in Epistemology". In: *Philosophical Studies* 67, pp. 133–50.

Feldman, Richard & Conee, Earl 2004: *Evidentialism: Essays in Epistemology*, Oxford: Clarendon Press.

Foley, Richard 1985: "What's Wrong with Reliabilism". In: *The Monist* 68, pp. 188–201.

Grundmann, Thomas & Misselhorn, Catrin 2003: "Transcendental Arguments and Realism". In: Hanjo Glock (ed.), *Kant and Strawson*, Oxford: Oxford University Press, pp. 205–18.

Grundmann, Thomas 2002: "Die Struktur des skeptischen Traumarguments". In: *Grazer Philosophische Studien* 64, pp. 57–81.

Grundmann, Thomas 2003: *Der Wahrheit auf der Spur: Eine Verteidigung des erkenntnistheoretischen Externalismus*, Paderborn: mentis.

Grundmann, Thomas 2007: "The Nature of Rational Intuitions and a Fresh Look at the Explanationist Objection". In: *Grazer Philosophische Studien* 74, pp. 69–87.

Grundmann, Thomas 2010: "Some Hope for Intuitions: A Reply to Weinberg". In: *Philosophical Psychology* 23, pp. 481–509.

Grundmann, Thomas 2011: "Defeasibility Theory". In: Sven Bernecker & Duncan Pritchard (eds.), *The Routledge Companion to Epistemology*, London, New York: Routledge, pp. 156–66.

Kornblith, Hilary 2002: *Knowledge and Its Place in Nature*, Oxford: Oxford University Press.

Lyons, Jack 2013: "Should Reliabilists be Worried about Demon Worlds?". In: *Philosophy and Phenomenological Research* 86, pp. 1–40.

McGinn, Colin 2004: *Mindsight*, Cambridge, MA: Harvard University Press.

Moore, G.E. 1959: "Certainty". In: G.E. Moore, *Philosophical Papers*, London: Allen and Unwin, pp. 227–51.

Pritchard, Duncan 2012: *Epistemological Disjunctivism*, Oxford: Oxford University Press.

Pryor, Jim 2000: "The Sceptic and the Dogmatist". In: *Nous* 34, pp. 517–49.

Sosa, Ernest 2007: "Dreams and Philosophy". In: E. Sosa, *A Virtue Epistemology*, Oxford: Oxford University Press, pp. 1–21.

Stroud, Barry 1984: *The Significance of Philosophical Scepticism*, Oxford: Clarendon Press.

Weatherson, Brian 2008: "Deontology and Descartes' Demon". In: *Journal of Philosophy* 105, pp. 540–69.

Weinberg, Jonathan, Nichols, Shaun & Stich, Stephen 2001: "Normativity and Epistemic Intuitions". In: *Philosophical Topics* 29, pp. 429–60.

Williamson, Timothy 2000: *Knowledge and Its Limits*, Oxford: Oxford University Press.

Wright, Crispin 1986: "Facts and Certainty". In: *Proceedings of the British Academy* 71, pp. 429–72.

Wright, Crispin 1991: "Scepticism and Dreaming: Imploding the Demon". In: *Mind* 100, pp. 87–116.

Wright, Crispin 2004: "Warrant for Nothing (and Foundations for Free)?". In: *Proceedings of the Aristotelian Society. Suppl. Vol.* 78, pp. 167–212.