The demon that makes us go mental: mentalism defended

Jonathan Egeland

Published online: 15 September 2018 © The Author(s) 2018

Abstract Facts about justification are not brute facts. They are epistemic facts that depend upon more fundamental non-epistemic facts. Internalists about justification often argue for mentalism, which claims that facts about justification supervene upon one’s non-factive mental states, using Lehrer and Cohen’s (Synthese 55(2):191–207, 1983) New Evil Demon Problem. The New Evil Demon Problem tells you to imagine yourself the victim of a Cartesian demon who deceives you about what the external world is like, and then asks whether you nevertheless have justification for your beliefs about the external world. Internalists and externalists agree that there is something that is epistemically good or valuable about both your actual beliefs and your beliefs in the demon scenario. Internalists claim that the epistemic property which these sets of beliefs share most intuitively should be thought of as sameness of justification. Externalists, on the other hand, reject this claim, usually either by challenging the internalist intuition directly, or by arguing that there is a more plausible way to think about the epistemic property in question. Recently, both kinds of externalist objection have been raised against the argument from the New Evil Demon Problem for mentalism. The goal of this paper is to defend the argument against three prominent objections—a pair of which is offered by Littlejohn (Can J Philos 39(3):399–434, 2009) and one by Williamson (in: Timmons M, Greco J, Mele A (eds.) Rationality and the good: critical essays on the ethics and epistemology of Robert Audi, Oxford University Press, Oxford, 2007; in: Dutant J, Dohrn D (eds.) The new evil demon, Oxford University Press, Oxford, 2016).

Keywords Epistemic justification · Epistemic internalism · New evil demon problem · Mentalism · Evidence

Jonathan Egeland
Jonathan.egeland.harouny@philosophy.su.se

Stockholm University, Springarvägen 5, 14261 Trängsund, Sweden

1
1 Mentalism

Facts about justification are not brute facts. They are epistemic facts that depend upon more fundamental non-epistemic facts. Internalists about justification often argue for mentalism, which claims that facts about justification depend upon one’s non-factive mental states.1,2 Externalists, on the other hand, reject this claim. In this paper, I will argue that the justification facts—i.e., the facts about which doxastic attitudes one now has justification (to a certain degree) to hold—depend upon one’s non-factive mental states in the following manner:

**Mentalism:** the justification facts supervene upon one’s non-factive mental states.

It is common to distinguish between three types of justification:

*Propositional justification:* the justification someone has for believing/withholding/disbelieving a certain proposition.

*Doxastic justification:* the justifiably held beliefs (or other doxastic attitudes) someone has.

*Personal justification:* someone’s being justified in believing/withholding/disbelieving a certain proposition.3

The focus of this paper will be on propositional justification, as my definition of the justification facts should make clear. So although certain formulations in this article might appear to be about personal or doxastic justification, it is propositional justification that is the topic at hand.4 Moreover, in contrast to the mentalist thesis above, I take facts about doxastic justification to be external. The reason is simply that doxastic justification plausibly can be defined as propositional justification plus proper basing,5 and that whether one satisfies the basing requirement is (at least in part) a causal, hence external, matter. So even though the externalist objections discussed in this article focus on doxastic justification, they aren’t levelled against the implausible view that proper basing somehow is an internalist notion, which means that the relevant issue really is the nature of propositional justification (i.e., whether it is internal or external).6

---

1 Proponents of mentalism include Pollock and Cruz (1999), Conee and Feldman (2001), Wedgwood (2002), Smithies (2012) and McCain (2016).

2 A mental state is non-factive just in case it doesn’t entail that it has propositional content that is true. By contrast, a factive mental state—like knowing that p—entails that its propositional content is true.

3 Cf. Littlejohn (2012, 5).

4 I suspect that it won’t always be possible to clearly differentiate talk about the various types of justification without additional commentary.

5 Turri (2010) provides counterexamples to the definition. Smithies (2015) responds that we simply can define proper basing as whatever turns propositional justification into doxastic justification. Thus, “immunity from counterexample may be gained at the cost of reduction” (Smithies 2015, footnote 19).

6 Some (e.g., Engel 1992; Littlejohn 2009, 2012) have argued that personal justification should be understood along internalist lines whereas doxastic justification should be understood along externalist lines. On this view, subjects in skeptical scenarios (like those that will be discussed in this paper) will be justified, whereas their beliefs won’t. However, not everyone agrees that personal and doxastic justification can come apart this way. For example, Kvanvig and Menzel (1990) argue that personal
Mentalism is often thought to receive support from our intuitive judgments about cases, most famous of which is the one described by Lehrer and Cohen’s (1983) New Evil Demon Problem. The New Evil Demon Problem tells you to imagine yourself the victim of a Cartesian demon who deceives you about what the external world is like, and then asks whether you nevertheless have justification for your beliefs about the external world. Internalists and externalists agree that there is something that is epistemically good or valuable about both your actual beliefs and your beliefs in the demon scenario. Internalists claim that the epistemic property which these sets of beliefs share most intuitively should be thought of as sameness of justification. Externalists, on the other hand, reject this claim, usually either by challenging the internalist intuition directly, or by arguing that there is a more plausible way to think about the epistemic property in question. Recently, both kinds of externalist objection have been raised against the argument from the New Evil Demon Problem for mentalism. In this paper, I will defend the argument against three of them—a pair of which attack the internalist intuition directly, and one that provides an alternative way of understanding the epistemic property in question.

This is the plan for the paper. Section 2 uses the New Evil Demon Problem to argue for mentalism. Section 3 defends this argument against three recent objections—two of which are offered by Clayton Littlejohn and one by Timothy Williamson. Section 4 concludes by taking stock and reflects on the relevance of the paper’s arguments for the current state of epistemology.

2 The new evil demon problem

The demon scenario, first presented by Lehrer and Cohen (1983; Cf. Cohen 1984), goes as follows:

The New Evil Demon Problem

Imagine that, unbeknown to you, you are the victim of a Cartesian demon who deceives you about what the external world is like. Although your beliefs about the external world are false, they’re supported by experiences that are, from your subjective point of view, indistinguishable from the experiences you would have if the world were exactly like you believe it to be.

Now, ask yourself whether your beliefs about the external world are justified despite being formed in ways that are unreliable. Intuitively, or so the argument goes, your beliefs in this scenario are just as justified as they would be if they were true—after all, it is conceivable that you’re actually in such a scenario right now. Thus, the New Evil Demon scenario indicates that changing the reliability of one’s doxastic dispositions doesn’t affect which propositions one now has justification to

Footnote 6 continued
justification entails doxastic justification; i.e., a person cannot be justified in believing that \( p \) while his belief that \( p \) isn’t justified.
believe. Instead, it seems that the facts about which propositions one now has justification to believe, and also the degree to which one has justification to believe them, remain unchanged as long as one holds fixed one’s non-factive mental states. Indeed, the best explanation (or so the internalist thinks) for why epistemic agents in normal cases and epistemic agents in subjectively indistinguishable demon cases seem to be equally justified in holding the same doxastic attitudes is that they have the same non-factive mental states. By abductive reasoning we can therefore conclude that mentalism is true: the justification facts supervene upon one’s non-factive mental states.

Mentalism provides a simple and natural explanation for shared philosophical intuitions about the New Evil Demon Problem. To see this, note that mentalism entails that if the non-factive mental states of epistemic agents are the same, then the facts about which propositions they now have justification to believe, and the degree to which they have justification to believe them, are also the same (Cf. Conee and Feldman 2001, 2). And since epistemic agents in normal cases and epistemic agents in subjectively indistinguishable demon cases have the same non-factive mental states, they are therefore equally justified in believing the same propositions.

Moreover, one need not consider extreme cases of global deception in order to appreciate this point. Consider the following case, which only involves local deception:

Illusory Ring

Claire is a collector of ancient artefacts. All the artefacts in her collection are bought from sources that seem highly trustworthy: she has vivid memorial beliefs that the sellers were honest and reliable. However, unbeknown to Claire, many of the sellers are notorious scam artists who, from time to time, sell replicas disguised as true originals. One of Claire’s artefacts – the rare ring with runic inscriptions from the Viking Age – is such a replica. One day, when Claire is looking through her collection of artefacts, she recalls her impression of the sellers and forms the belief that the ring with runic inscriptions is from the Viking Age.

---

7 Originally, the scenario was presented as a counterexample to the claim that reliability is necessary for justification. Other counterexamples to this claim are provided by Putnam’s (1981) envatment scenario and, more popularly, the Matrix movies. On the other hand, counterexamples to the claim that reliability is sufficient for justification are provided by BonJour (1985, 41), Lehrer (1990, 163–164), Plantinga (1993, 199) and Smithies (2014).

8 A variety of reliabilist responses have been offered. Goldman (1986, 1988), Comesan˜a (2002), Sosa (2003) and Majors and Sawyer (2005) try to come up with a reliabilist view that accommodates the intuition elicited by the New Evil Demon Problem. On the other hand, Bach (1985), Brewer (1997), Engel (1992), Sutton (2007) and Lyons (2013) claim that they don’t share the internalist intuition or that it simply is wrong. For discussions of many of these objections, see Fumerton (1995, Ch. 4); and Littlejohn (2012, Introduction).

9 Wedgwood (2002) similarly argues that the intuitions that underwrite the New Evil Demon Problem for reliabilism also underwrite an argument against all versions of externalism about justification.

10 I operate with a distinction between having a memory/memorial belief, which is a non-factive mental state, and remembering that something is the case, which is a factive mental state.
Although Claire’s belief in the Illusory Ring case is unreliable and false, it doesn’t seem any less justified than her belief would be in a subjectively indistinguishable case where the sellers actually are trustworthy and the ring is from the Viking Age. Once again, the best explanation for why Claire’s beliefs are equally justified in both cases is that she has the same non-factive mental states in both. As long as one’s non-factive mental states are held fixed, there can be no change in which doxastic attitudes one now has justification to hold or the degree to which one has justification to hold them.

It is often assumed in contemporary epistemology that the justification facts are a function of one’s (epistemic) reasons for holding certain doxastic attitudes. In the discussions that follow, I will rely on this assumption when I talk about reasons for belief (or other doxastic attitudes). And, with this assumption in place, mentalism can plausibly be interpreted as a thesis about what reasons one has. Given that the justification facts are a function of one’s reasons, a plausible construal of mentalism says that one’s reasons are provided by one’s non-factive mental states. Alternatively, adopting the evidentialist’s parlance, we can say that one’s evidence is provided by one’s non-factive mental states. Thus, insofar as concepts like reason or evidence are used in this article, this is what the internalist will be taken to be committed to.

However, cases like these and the intuitive epistemological lessons they are supposed to teach haven’t gone unchallenged. Lately, the argument from the New Evil Demon Problem for mentalism has come under heavy fire. In the next section, I will defend the argument against three objections—a pair of which is offered by Littlejohn and one by Williamson.

---

11 I take it that in order for a belief-forming process to be reliable it must produce a sufficient amount of true beliefs both locally (i.e., in scenarios closely linked to the case under consideration) and globally (i.e., in different scenarios in a wide range of cases). For a discussion of local and global reliability, see Craig (1991).

12 For a discussion by someone who holds a dissenting view, see Broome (2013).

13 Plausibly, they’re either identical to a subset of one’s non-factive mental states or their propositional content.

14 After the publication of Williamson’s Knowledge and Its Limits (2000), it has become common for epistemologists to think that only true propositions (or facts) can be evidence. (Williamson (2000, 185), of course, famously argues for the $E = K$ Thesis: “knowledge, and only knowledge, constitutes evidence”, and later (2007, 120) for what we may call the $R = K$ Thesis: “that $p$ is available to you as a reason if and only if you know $p$.” Cf. Hyman (2006) who also endorses the $E = K$ thesis, but arrives at it via another route than Williamson). However, this is not a commitment that we need to make. Indeed, there is also an older tradition preceding Williamson that understands evidence in terms of what goes on inside the head rather than outside it, and that allows false propositions (like the content of false visual experiences) to be included in one’s evidence. Here, I simply want to point out that the internalist naturally aligns himself with something more akin to the traditional view. For a useful survey of some of the ways in which the concept of evidence has been understood, see Kelly (2008).

15 Here, I will remain neutral on whether one’s reasons/evidence is propositionally specified. For useful (and disagreeing) discussions of the issue, see Turri (2009) and Glüer and Wikforss (2018).
3 Objections and replies

3.1 Littlejohn’s first objection

According to the first objection offered by Littlejohn, our intuitions about the New Evil Demon Problem support the following thesis:

*Parity*: Necessarily, if S and S′ are epistemic counterparts [which is to say that they have the same non-factive mental states], S is justified in believing p iff S′ is justified in believing p. (Littlejohn 2009, 400.)

However, as Littlejohn points out, Parity doesn’t entail mentalism since epistemic counterparts can be justified in believing the same propositions, but without having the same reasons/evidence. For example, epistemic counterparts may satisfy Parity and thus be justified in holding the same beliefs, while one of them has external reasons/evidence that the other doesn’t have. Moreover, as a result, Parity also allows that there can be (external) factors that make the epistemic agent in the normal case more justified in holding certain beliefs than his counterpart in the bad case is. Indeed, for all that Parity tells us, epistemic counterparts are always justified in holding the same beliefs. But it doesn’t tell us anything about what their reasons/evidence is or the degree to which they are justified in holding the same beliefs; perhaps there are external reasons that play a justification-conferring role and that make the epistemic agent in the normal case more justified in holding the beliefs he does.

In order to rule out such a position, one must also endorse a stronger thesis:

*Equality*: Necessarily, if S and S′ are epistemic counterparts, the fact that q is relevant to the justification of what S believes iff the fact that q is relevant to the justification of what S′ believes. (Littlejohn 2009, 406.)

However, Littlejohn claims that Equality has implausible consequences and therefore is vulnerable to a modus tollens.16 To see why this is so, recall the case about the Illusory Ring from the previous section. Consider Claire and an epistemic counterpart, Blaire, both of who believe that the ring with the runic inscriptions is from the Viking Age. The only difference is that in Claire’s case the sellers are untrustworthy and the ring is a perfect replica, whereas in Blaire’s case the sellers are trustworthy and the ring is a true original. Now, let’s say that their reason for

---

16 Another interesting view that clearly rejects Equality is epistemological disjunctivism. (See, e.g., McDowell (1982), Snowdon (1981–1982, 1990) and Pritchard (2012).) According to disjunctivists, the perceptual experiences (especially visual experiences) of epistemic counterparts in normal and bad cases are fundamentally different both in kind (see McDowell 1982, 472ff.) and insofar as the agent in the normal case can have justification or rationality-conferring reasons that the counterpart in the bad case doesn’t have (see Pritchard 2012, 16, Cf. 2012, 42). Moreover, it is interesting to note that despite his rejection of Equality, Pritchard (2012, 1) claims that epistemological disjunctivism is the “holy grail” of epistemology because it is able to reconcile internalism (with its demand for an accessibility condition on justification) and externalism (with its demand for a strong connection between justification and truth). However, in a recent paper Madison (2014) argues that by failing to do justice to internalist intuitions and the theses they support (like Equality), disjunctivists are ultimately unable to reconcile internalism and externalism.
believing that the ring is from the Viking Age is that they both have a memory to the effect that they bought it from a trustworthy source.\textsuperscript{17} Blaire’s reason is her memorial knowledge that the seller is a trustworthy source, whereas Claire’s reason is her false memorial belief that the seller is a trustworthy source.\textsuperscript{18} Plausibly, Littlejohn tells us, knowledge of a proposition $p$ is sufficient for $p$’s inclusion in one’s evidence; so the proposition that the seller is trustworthy is included in Blaire’s evidence and rationally supports her belief that the ring is from the Viking Age. On the other hand, the falsity of a proposition $p$ is surely sufficient for the exclusion of $p$ from one’s evidence; so the proposition that the seller is trustworthy is not included in Claire’s evidence and doesn’t rationally support her belief that the ring is from the Viking Age.\textsuperscript{19} However, if Equality is true this cannot be the case since it tells us that there can be no difference in the facts that are relevant to the justification of the beliefs of epistemic counterparts like Claire and Blaire. Equality therefore has the implausible consequence that Blaire’s memorial knowledge either need not be seen as providing a true proposition that is included in her evidence, or Claire’s memorial belief can be seen as providing a false proposition that is included in her evidence. Littlejohn puts the issue in the form of a trilemma:

To save Equality, you either have to say that we cannot have knowledge of propositions about the external world, that there can be false propositions included in someone’s evidence, or that knowledge is not enough for a proposition’s inclusion in someone’s evidence. (Littlejohn 2009, 407.)

Now, there are two plausible ways of responding to this particular objection. First, one can accept the second of Littlejohn’s triad of claims by affirming that there indeed can be false propositions included in someone’s evidence. In the Illusory Ring case we saw that, intuitively, Claire’s belief about the origins of the ring was just as justified as it would have been if she was in Blaire’s epistemic position—i.e., if the sellers actually were trustworthy and the ring in fact was from the Viking Age. And since Blaire’s memorial knowledge that the seller is trustworthy clearly provides evidence for her belief in the origins of the ring, so must Claire’s false belief.\textsuperscript{20,21}

\textsuperscript{17} Their memory should in this instance be understood as a \textit{motivating reason}, rather than a \textit{normative reason}, so as not to beg any questions against the externalist. For more on this distinction, see McNaughton and Rawling (2018).

\textsuperscript{18} This is to say that Blaire \textit{remembers} that the seller is trustworthy, whereas Claire only has a \textit{memory} that the seller is trustworthy.

\textsuperscript{19} In other words, whereas Claire’s memory only is a motivating reason, Blaire’s memory also provides a normative reason.

\textsuperscript{20} Counterexamples to the claim that false propositions cannot be included in one’s evidence abound. See, e.g., Warfield (2005), Fitelson (2010), Klein (2008), Arnold (2011) and Rizzeri (2011). Littlejohn (2012, 106–109) argues against this kind of example, and McCain (2016, 25–27) responds to his argument.

\textsuperscript{21} In other words, the internalist need not accept the Williamsonian idea, with which Littlejohn is sympathetic (although he (2011) rejects the $E = K$ Thesis), that only true propositions can be evidence. Focusing only on the objection at hand, the present case about Blaire and Claire doesn’t carry much force.
There is, undeniably, an epistemic difference between Claire and Blaire’s beliefs about the trustworthiness of the seller. Whereas Claire’s belief, which seems to be true from her subjective point of view, in fact is false, Blaire’s belief is true and satisfies all the (externalist) conditions for knowledge. The point is just that not all epistemic differences are justificationally relevant. What is relevant for the inclusion one’s mental states or their propositional content in one’s evidence is that they satisfy non-factive epistemic conditions. For example, on a plausible view of knowledge, one needs an unGettiered (doxastically) justified true belief in order to know that something is the case. Blaire therefore satisfies these epistemic conditions. But what is relevant for the inclusion of her memorial knowledge or its propositional content that the person who sold her the ring is a trustworthy source in her evidence is that the mental state (or its propositional content) satisfies non-factive epistemic conditions—like being a belief (or the propositional content thereof). In general, if a mental state/proposition involves elements (non-factive epistemic conditions) that are sufficient for explaining certain effects (a change in one’s stock of evidence) which that state/proposition produces, then it is those elements that really are responsible for those effects. Blaire’s memorial knowledge (or its propositional content) is therefore included in her evidence, but

Footnote 21 continued
against the internalist. Indeed, our intuitions seem to favor the view that one’s evidence is, as suggested above, provided by one’s non-factive mental states and thus that false propositions can be included.

There are, of course, other considerations that motivate Williamson and Littlejohn’s views on evidence. For example, Williamson (2000, 193) argues that in normal scientific discourse, evidence is treated as true propositions that are commonly known or perhaps just widely available in the scientific community. I think this is right. However, I suspect that if the scientist or scientific community is pressed to specify what they take their evidence to be, in light of certain possible skeptical scenarios, by citing what their evidence/reasons are for various scientific hypotheses and theories in a variety of cases, that they will find themselves referring to certain non-factive mental states or their propositional content—perhaps they will endorse what may be called the E = B Thesis: that only propositions believed or accepted by the scientific community are evidence. For a critical, and much more detailed discussion of Williamson’s views, see McGlynn (2014).

An anonymous referee pointed out that the proponent of the view that one’s evidence is propositionally specified and can include false propositions will have a hard time maintaining consistency, which says that one’s propositionally specified evidence must be consistent (and which is motivated by the way in which conditional probabilities are treated in probability theory). In response, the proponent of this kind of view can always pursue the idea that even though false propositions can be included in one’s evidence, not all of them can—in particular, he may say that inconsistent propositions never are included. However, unless he is able to motivate this idea, the suggestion appears to be ad hoc.

Although this raises issues that are somewhat beyond the purposes of this article, I think one way to motivate the idea is to consider various possible scenarios involving (what appears to be) false evidence and see whether they can teach us something about which (if any) constraints there are on false evidence. Doing this, I wouldn’t be surprised if we found ourselves convinced that false propositions often can be evidence, but that inconsistent propositions never can be evidence.

Some epistemologists have rejected this view. For example, Foley (1993) and Alston (1989) deny that justification is necessary for knowledge, whereas Sutton (2007) claims that justification is sufficient for knowledge.

Contrary to Williamson’s knowledge first approach, this presupposes that knowledge can be factored into constituent parts.

The same point is also made by Wedgwood (2002, 362–363).
only under the guise of being a mental state (or proposition) that satisfies certain non-factive epistemic conditions. And since Blaire and Claire are stipulated to satisfy the same non-factive epistemic conditions, they therefore have the same evidence and, hence, equally justified beliefs about the origins of the ring.

Second, claiming that our (including the internalist’s) intuitions about the New Evil Demon Problem only support Parity and not Equality is question-begging. As I formulated the argument in the previous section, the New Evil Demon Problem supports the mentalist thesis according to which the justification facts supervene upon one’s non-factive mental states. And formulated in this manner, mentalism entails Equality. The intuitions appealed to in my presentation of the argument from the New Evil Demon Problem therefore do support Equality and, moreover, were also thought to do so from the very beginning. According to Cohen’s (1984, 281) original version, “… on the demon hypothesis, we would have every reason for holding our beliefs, that we have in the actual world.”26 So even though I take the first reply to be successful, Littlejohn’s objection doesn’t really have what it takes to get off the ground. Instead, it simply assumes that the argument from the New Evil Demon Problem is false by claiming that our intuitions don’t support Equality. In order for this claim to be able to do its intended work, Littlejohn would have to argue that our intuitions about the New Evil Demon Problem somehow fail to support Equality—and this brings us to his second objection.

3.2 Littlejohn’s second objection

According to Littlejohn’s second objection,27 internalist intuitions about the New Evil Demon Problem are undermined by other intuitions that support the denial of mentalism. Consider the following thesis:

Asymmetry: It is possible for there to be a pair of epistemic counterparts, S and S’, such that (a) only one of the pairs has good enough reason to believe p and have a justified belief about p or (b) S has better reasons to believe p than S’ does. (Littlejohn 2009, 408.)

Littlejohn claims that Asymmetry receives intuitive support from reflection on a case involving an epistemic agent, Alice, who imagines her counterpart, Cooper, who is deceived by a Cartesian demon. In this case, Alice should believe the following two claims:

(1) If there is no more reason for me to believe I have hands than there is for Coop to believe he has hands, I should not believe I have them. (Littlejohn 2009, 408.)

---

26 That the new evil demon problem supports mentalism and, as consequence, Equality is also argued by Wedgwood (2002) and Smithies (forthcoming).

27 What I’m here calling Littlejohn’s second objection is, as will become clear, really two objections.
(2) If my reasons for believing I have hands are no better than his, I really have no good reason for believing I have hands. (Littlejohn 2009, 408.)

And she should reject the third:

(3) Even if there is no more reason for me to believe I have hands than there is for Coop to believe he has hands, there is nothing wrong with my continuing to believe I have hands. (Littlejohn 2009, 408.)

Now, if we supply the case with a little more detail, I think most philosophers will discover that their intuitions actually speak in favor of (a slightly modified and internalist version of) the third claim and neither of the former two. Consider the first claim. The reason Alice shouldn’t believe she has hands, on the assumption that she has no more reason to believe it than Coop, is that Coop’s non-factive mental states don’t provide him with sufficient reason to believe that he has hands. But does it seem plausible that Coop doesn’t have sufficient reason to believe that he has hands? Coop has visual experiences of having hands, experiences of other people telling him that they see him having hands, and it clearly appears as if he is successfully interacting with his environment by using his hands. Indeed, from Coop’s perspective, it clearly seems that he should believe that he has hands because he has several good reasons for believing that he does. And this fact—namely, that his belief that he has hands is strongly supported by his non-factive mental states—indicates that he does have sufficient reason to justifiably hold his belief.

When it comes to the second claim, the reason Littlejohn thinks that Alice has no good reason to believe she has hands, on the assumption that her reasons are no better than Coop’s, is that Coop’s non-factive mental states don’t provide him with a single good reason to believe that he has hands. However, Coop has several kinds experience (visual, testimonial, tactile, etc.) supporting his belief that he has hands. And, once again, having visual experiences of having hands, auditory experiences of other people talking about one’s having hand, and feeling and using what clearly appears to be one’s hands on a daily basis seems to provide one with not just one, but several good reasons for believing that one has hands. Intuitively, it therefore seems that Coop does have good reason to believe that he has hands.

Moreover, this is of course just what the New Evil Demon Problem indicates. Our experiences and beliefs provide us with good reasons for holding beliefs about the external world, despite their fallibility and unreliability. And these intuitions also support the third of Littlejohn’s claims. The reason that there is something wrong with Alice’s belief that she has hands, on the assumption that she has the same reasons as Coop, is that there also is something wrong with Coop’s belief that

---

28 Littlejohn also mentions two other claims. But since they are very similar to (1) and (2)—only given a third-person formulation—I’ll assume that it suffices to respond to the latter ones.

29 It is actually quite hard to see whether Littlejohn’s case is supposed to differ in any significant way from the original New Evil Demon scenario. Does the fact that we now are supposed to evaluate someone (Alice) who imagines someone else (Cooper) in a demon scenario, instead of simply evaluating someone in a demon scenario, really change matters? I have a hard time seeing why it should.
he has hands. Now, epistemically, there is of course something wrong with Coop’s belief—even the most ardent internalists should agree. For example, it is neither reliable nor true. However, his belief clearly seems justified. With all the aforementioned experiences to back it, his belief cannot be faulted for any rational failure. On the basis of these considerations, I therefore conclude that (3) is supported by intuition whereas (1) and (2) isn’t.

However, it is always possible for Littlejohn to insist that he has intuitions to the contrary and that they have to be respected. But seeing as most professional philosophers appear to have internalist intuitions—at least judging by the literature surrounding scenarios like the ones we’ve considered above—it seems safe to say, at least for now, that mentalism on balance receives more support from intuition than externalism.

Another worry Littlejohn has about internalist intuitions is that empirical research has shown that when it comes to action people usually think about justification in externalist terms, and that one should not expect people to have different intuitions about the justification of belief. The reason one shouldn’t adopt a “hybrid view” which is internalist about the justification of belief and externalist about the justification of action is that anyone committed to such a view has to reject the following plausible principle:

**Link**: If S’s belief that \( U \)-ing is permissible is justified, S’s \( U \)-ing is justified.

(Littlejohn, 2009, 411.)

And that by rejecting Link, one also has to reject another intuitively plausible principle:

---

30 I suspect this is the real reason why Littlejohn thinks that the Alice and Coop scenario supports Asymmetry. However, the notion of epistemic wrongness, which (3) appeals to, needs to be unpacked. Without specifying what epistemic wrongness consists of, it cannot be considered a sufficient condition of having an unjustified belief that the belief somehow is wrong.

A natural way of interpreting epistemic wrongness is to say that a belief is wrong if and only if it is false. But it is clearly possible to have justified false beliefs. Of course, no philosophical view is immune from dissent—Littlejohn himself being a case in point. However, the idea that justification entails truth is undeniably considered quite implausible by the majority of contemporary epistemologists, with most of them finding it counterintuitive or unmotivated.

31 Turri (2015a, b) has recently provided empirical evidence that ordinary belief evaluation by non-philosophers tend to support externalism. However, in philosophical matters it seems reasonable that expert intuition should be given more authority. After all, I take it that few (if any) philosophers think it is reasonable to consider justified true belief sufficient for knowledge just because ordinary evaluation by non-philosophers might support a pre-Gettier conception of knowledge.

32 An anonymous referee objected that my reasoning might involve a fallacious appeal to authority, and that if it doesn’t, it relies on unpopular assumptions. In response, I will simply point out that appealing to authority X isn’t fallacious as long as we have good reason to think that X is a reliable source of information when it comes to the subject matter at hand. Moreover, when it comes to technical epistemological issues, like those that are the topic of this article, we should only rely on the philosopher’s intuitions insofar as we have reason to think that they are a reliable source of information about those issues. And, indeed, this appears to be an assumption underlying standard philosophical practice; the whole dialectic of analysis and counterexample seems to be a case in point. Also, there is a growing literature supporting the claim that intuitions (or their propositional content) constitute evidence. For interesting discussions, see Bengson (2014), Devitt (2015) and Koksvik (2017).
Fault: If S’s Φ-ing were unjustified, S could be faulted for having Φ’d. (Littlejohn, 2009, 412.)

There are different ways of responding to this worry. For example, one can endorse the “hybrid view” and claim that rejecting Fault isn’t too bad after all. However, I think a better option is to argue that Littlejohn has presented us with a false dilemma since the proponent of the “hybrid view” only has to reject Fault on the not-so-plausible assumption that a person cannot be faulted for actions committed on the basis their justified beliefs. And this is something he acknowledges: “Surely if someone was not wrong to believe that they ought to or are permitted to Φ and Φ’d from the right sort of motives, they could not be faulted for having Φ’d.” (Littlejohn, 2009, 412.) But it is not at all obvious that the standards for praise and blame are the same when it comes to action and belief. Consider the following example:

Indoctrination Victims
Jane is a victim of indoctrination. All her life she has been provided one-sided evidence that all and only people who commit suicide before the age of 18 enter into Paradise—e.g., trustworthy people whom she knows to be reliable about metaphysical and religious matters, but who only happen to be wrong about this particular issue, have told her so for as long as she can remember. As a result, Jane believes that her only chance of entering paradise is to commit suicide before she turns 18. And motivated by her belief (and her desire to enter into paradise), tragically enough, she does. Similarly, Twin Jane, an epistemic counterpart of Jane, is also an indoctrination victim. Just like with Jane, Twin Jane has been provided one-sided evidence all her life that all and only people who commit suicide before the age of 18 enter into Paradise by the same kind of reliable testifiers. As a result, Twin Jane believes that her only chance of entering paradise is to commit suicide before she turns 18. And motivated by her belief (and her desire to enter into paradise), she does. However, in the case of Twin Jane, her indoctrinators are actually right. Moreover, it is common knowledge among people that one will in fact enter paradise just in case one commits suicide before turning 18.

Now, it is clear from the description of the example that both Jane and Twin Jane’s beliefs are justified. Their evidence, one-sided though it is, supports their beliefs. However, in the case Jane, it clearly seems that her action can be faulted. There are obvious reasons for why she shouldn’t have committed suicide. For example, by doing so she will not in fact enter into paradise, she will “miss out” on good experiences she could have had, and her loved ones will probably have to endure a lot of unnecessary grief and suffering as a result. For these and similar reasons, it therefore seems completely natural to say that she shouldn’t have committed suicide even though her evidence supported believing that she should.

---

33 Not only does this seem intuitively right, it is also predicted by plausible internalist and externalist (e.g., reliabilist) views.
On the other hand, it doesn’t seem that Twin Jane’s action can be faulted. Indeed, the reasons for acting as she does far outweigh the reasons for not doing so. For example, as a result of committing suicide before she turns 18, Twin Jane doesn’t only enter paradise, she also doesn’t miss out on any good experiences (her experiences in paradise will be far better), and her family and loved ones will not suffer the tiniest bit (let’s stipulate that they’ve themselves already committed suicide or that they’re just genuinely happy for the fact that Twin Jane is in a better place). The general point is that it is possible to separate our reactive attitudes toward an action from our reactive attitudes toward the belief(s) which (together with the appropriate desires) motivated it. Hence, as the case above indicates, a person can fault an action committed on the basis of their justified belief(s).34

Moreover, the internalist about the justification of belief can also avoid the objection by committing himself to internalism about the justification of action (which I am very sympathetic to). For example, he might do this by arguing (as I similarly did above) that even though most people have externalist intuitions about the justification of action, when it comes to professional philosophers things are different, and that expert intuition should be given more weight.35 Doing this, the internalist need not argue against Littlejohn’s assumption that a person cannot be faulted for an action committed on the basis of their justified belief(s). However, he can still do so. In that case, he should say that Jane’s (but not Twin Jane’s) action is blameworthy but justified.

3.3 Williamson’s objection

If the argument above is sound, then justification and blamelessness can come apart. But might not this create other problems for internalism? Perhaps one can object to

34 My suggestion of course goes against the widely shared assumption that the culpability (assuming this is what Littlejohn means by being subject to “fault”) of action is to be understood along internalist lines. Although undoubtedly controversial, I think the example above provides a plausible case for the opposite view. Another relevant example is the following:

Nuclear Attack

Imagine that the President of a mighty nation launches an atomic bomb against another nation (let’s call it “X”) that has been hostile and warmongering for decades after being persuaded by his military advisors that X is close to developing their own nuclear weapons, and that unless he launches a preemptive nuclear attack X will likely terrorize and bomb the rest of the world with their own nuclear weaponry. As it turns out, however, the President’s military advisors are wrong. X is not developing nuclear weapons. In fact, its leaders have recently decided, after reading Kant’s *Groundwork of the Metaphysics of Morals*, to live and govern in accordance with the categorical imperative.

Now, there is clearly a sense in which the President shouldn’t have bombed X even though he was justified in believing that he should. Indeed, it will in fact cause many deaths and extreme amounts of unnecessary suffering. And, for those reasons, one may think that the President does deserve (at least some) blame for his action.

35 Littlejohn (2012, Ch. 6) raises other objections against internalism about the justification of action, and in order to properly defend that kind of view those objections have to be met. However, the point I’m presently making is simply that the argument hinted at above provides an answer to the particular objection under consideration, rather than a full-fledged defense of internalism in the realm of action.
the argument from the New Evil Demon Problem by claiming that the beliefs had by epistemic agents in demon cases are blamelessly unjustified—and that is precisely what Williamson does. According to the third objection, justification and blamelessness can come apart, and demon cases are prime examples where this is the case. Although the beliefs had by epistemic agents in demon cases are unjustified—presumably because they either are unreliable or fail to constitute knowledge—they nevertheless have an excuse for holding their beliefs and therefore cannot be blamed. This is how Williamson puts it:

…the victim of a paradigmatic skeptical scenario is not to be blamed for forming false beliefs under the misapprehension that they constitute knowledge. The subject has a cast-iron excuse for having formed those beliefs…[However,] excusable failure is not normatively equivalent to success. (Williamson 2007, 116–117.)

So, in other words, whereas the epistemic agent in the normal case is blameless and justified, the counterpart in the demon case is blamelessly unjustified.

However, the objection fails to recognize the way in which our intuitions about justification are sensitive to a distinction between perceptual failings and cognitive failings. In order to illustrate this, Pryor (2001, 117) has us consider three victims of skeptical scenarios. The first victim is neither blameless nor justified: he fails to respect his perceptual evidence and simply believes whatever he feels like believing. The second victim is blameless but unjustified: he tries his best to form beliefs that are supported by his perceptual evidence, but fails due to brainwashing or some reason-distorting drug he’s been given. The third victim is both blameless and justified: he successfully forms the beliefs that are supported by his perceptual evidence. By claiming that victims who respect or try to respect their perceptual evidence when forming beliefs about the world simply are blamelessly unjustified—like the second and third victims above—Williamson’s proposal collapses an intuitive distinction between justified perceptual failings and unjustified cognitive failings.

Recently, Williamson (2016) has responded to this worry by distinguishing the act of following a norm and merely being disposed to follow a norm. His claim is that while victims of skeptical scenarios are unable to follow the norms of justification, they may be disposed to do so—in which case they have an excuse for their beliefs which they wouldn’t have had if they weren’t so disposed. The upshot is that a victim’s beliefs are only really blameless (or excusable) when formed as a
result of a disposition to follow norms that normally would produce justified beliefs.  

However, the problem with this response is that it doesn’t really account for the way in which our intuitions about justification are sensitive to the difference between perceptual failings and cognitive failings. To see why this is so, consider Pryor’s second and third victims again. Both of them are generally disposed to form justified beliefs, which means they are blamelessly unjustified. But, intuitively, an obvious justificatory difference still persists between them: it is only the third victim which forms his beliefs on the basis of his perceptual evidence. Now, Williamson might claim that the brainwashing/reason-distorting drug somehow changes the dispositions of the second victim so that he no longer is disposed to comply with the norms for justification. But that wouldn’t help to clarify the intuitive difference between the victims. For if the second victim neither is blameless nor justified, then he seems to be in the same epistemic position as the first victim. But that would be absurd: simply believing whatever one feels like, without any effort to respect one’s evidence, is clearly more blameworthy than doing one’s best in forming beliefs that conform to the evidence one has. Moreover, one can also come up with similar cases where the victim’s dispositions are guaranteed to remain uncompromised, in which case the intuitive difference between that victim and Pryor’s third victim still goes unaccounted for. In fact, Williamson himself provides such a case:

Brain Scrambler
A special device emits waves of some sort with a selective scrambling effect on brains. The waves inflict no permanent damage, and do not even change what ‘programme’ the brain is running, but they occasionally alter the contents of unconscious short-term working memory, so that some computations

40 “But once we realize how easily norms for belief which one violates in sceptical scenarios generate secondary and tertiary norms [i.e., higher order norms about one’s dispositions to follow the lower order norms] with which one complies in those scenarios, we should abandon the naïve idea that the normative status which the subject’s beliefs share in the good and bad cases might be justification, as opposed to blamelessness.” (Williamson 2016, 22).

41 Another response offered by Littlejohn (forthcoming) is that internalist arguments involving skeptical scenarios, like Pryor’s three victims or the New Evil Demon Problem, are false dilemmas (or, as Littlejohn calls them, “Contrast Arguments”). As he sees it, they involve the following pattern of reasoning:

1. The beliefs of the victim of a skeptical scenario have something positive going for them.
2. The victim’s beliefs are either blameless or justified.
3. They are not merely blameless.
4. Therefore, they must be justified. (See Littlejohn forthcoming, 7–8.)

But, Littlejohn (forthcoming, 8–11) claims by following Strawson (1962) and others, the victim’s beliefs may satisfy other positive epistemic properties, such as being excusable or exemptible. However, Madison (2017, 9–11) has recently pointed out a plausible line of response for the internalist. According to Madison, the internalist can replace the third premise with a more comprehensive one saying that the victim’s beliefs are not simply blameless, or excusable, or exemptible, etc. This way, the fact that the victim’s beliefs may satisfy other epistemic properties besides justification or blamelessness doesn’t threaten the conclusion by rendering the argument fallacious.

42 Cf. Smithies (forthcoming, Ch. 2).
produce incorrect results. Thus a normal subject may confidently announce that $17 + 29 = 33$. Similarly, consider Connie, a normal subject and competent mathematician who sincerely and confidently announces that 179 is and is not prime, because a scrambled piece of reasoning yields that conclusion, and a scrambled application of a contradiction-detector failed to sound the alarm in retrospect. (Williamson 2016, 14.)

As Williamson makes clear, Connie is disposed to form justified beliefs about mathematics and confidently exercises that disposition. Nevertheless, since her beliefs are the result of a brain-scrambling machine that makes her prone to cognitive failings, they are, just like those of Pryor’s second victim, blamelessly unjustified. Williamson response therefore fails to account for the intuitive distinction between justified perceptual failings and unjustified cognitive failings.

4 Concluding remarks

Let me conclude by taking stock. In Sect. 2, I used Lehrer and Cohen’s New Evil Demon Problem to argue for mentalism. In Sect. 3, I defended the argument against three recent objections—a pair of which is offered by Littlejohn and one by Williamson.

Mentalism is an intuitively compelling thesis about how facts about justification depend upon more fundamental non-epistemic facts. By assessing the justificatory status of epistemic agents in various skeptical scenarios, we can infer that the justification facts supervene upon one’s non-factive mental states. However, this kind of argument from cases has recently come under heavy fire by proponents of a factive turn in epistemology. Philosophers like Littlejohn and Williamson argue that the intuitions supporting internalist views of justification can be undermined, and that in doing so they pave the way for an alternative kind of view according to which facts about justification depend upon facts about the external world, factive mental states, or other externalist conditions. The goal of this paper has been to show that the objections raised by Littlejohn and Williamson fail to undermine the intuitions that support internalism about justification. Although the factive turn has redirected much of contemporary epistemology away from its Cartesian roots, I conclude that a volte-face toward the traditional non-factive paradigm is warranted.

Acknowledgements For helpful discussions and comments, I want to thank Åsa Wikforss, Mikael Janvid, and Duncan Pritchard.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Funding Work on this paper has received funding from the European Union’s Horizon 2020 Research and Innovation programme under Grant Agreement no. 675415.
References

Alston, W. P. (1989). Epistemic justification: Essays in the theory of knowledge. Ithaca and London: Cornell University Press.

Arnold, A. (2011). Some evidence is false. Australasian Journal of Philosophy, 91, 165–172.

Bach, K. (1985). A rationale for reliabilism. The Monist, 68, 246–263.

Bengson, J. (2014). How philosophers use intuition and ‘Intuition’. Philosophical Studies, 171(3), 555–576.

BonJour, L. (1985). The structure of empirical knowledge. Cambridge: Harvard University Press.

Brewer, B. (1997). Foundations of perceptual knowledge. American Philosophical Quarterly, 34, 41–55.

Broome, J. (2013). Rationality through reasoning. Oxford: Blackwell.

Chisholm, R. (1989). Theory of knowledge. Englewood Cliffs: Prentice-Hall Inc.

Cohen, S. (1984). Justification and truth. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 46(3), 279–295.

Comesaña, J. (2002). The diagonal and the demon. Philosophical Studies, 110, 249–266.

Connee, E., & Feldman, R. (2001). Internalism defended. American Philosophical Quarterly, 38(1), 1–18.

Craig, E. (1991). Knowledge and the state of nature: An essay on in conceptual synthesis. Oxford: Clarendon Press.

Devitt, M. (2015). Relying on intuitions: Where Cappelen and Deutsch go wrong. Inquiry: An Interdisciplinary Journal of Philosophy, 58(7–8), 669–699.

Engel, M. (1992). Personal and doxastic justification. Philosophical Studies, 67, 133–151.

Fetelson, B. (2010). Strengthening the case for knowledge from falsehood. Analysis, 70, 666–669.

Foley, R. (1993). Working without a net: A study of egocentric epistemology. New York, Oxford: Oxford University Press.

Fumerton, R. (1995). Metaepistemology and skepticism. Lanham: Rowman & Littlefield Publishers Inc.

Ginet, K. (1975). Knowledge, perception, and memory. Dordrecht: D. Reidel.

Glu¨ er, K., & Wikforss, A˚. (2018). Reasons for belief and normativity. In D. Star (Ed.), Oxford handbook of reasons and normativity. Oxford: Oxford University Press.

Goldman, A. (1986). Epistemology and cognition. Cambridge: Harvard University Press.

Goldman, A. (1988). Strong and weak justification. Philosophical Perspectives, 2, 51–69.

Hyman, J. (2006). Knowledge and evidence. Mind, 115(460), 891–916.

Kelly, T. (2008). Evidence: Fundamental concepts and the phenomenal conception. Philosophy Compass, 3(5), 933–955.

Klein, P. (2008). Useful false beliefs. In Q. Smith (Ed.), Epistemology: New essays (pp. 25–61). Oxford: Oxford University Press.

Kocksvik, O. (2017). The phenomenology of intuition. Philosophy Compass, 12(1), e12387.

Kvanvig, J., & Menzel, C. (1990). The basic notion of justification. Philosophical Studies, 59, 235–261.

Lehrer, K. (1990). Theory of knowledge. Boulder: Westview Press.

Lehrer, K., & Cohen, S. (1983). Justification, truth, and coherence. Synthese, 55(2), 191–207.

Littlejohn, C. (2009). The externalist’s demon. Canadian Journal of Philosophy, 39(3), 399–434.

Littlejohn, C. (2011). Evidence and knowledge. Erkenntnis, 74(2), 241–262.

Littlejohn, C. (2012). Justification and the truth-connection. Cambridge: Cambridge University Press.

Littlejohn, C. (forthcoming). A Plea for Epistemic Excuses. In: Dorsch, F., & Dutant, J. (eds.), The new evil demon problem. Oxford University Press.

Lyons, J. (2013). Should reliabilists be worried about demon worlds? Philosophy and Phenomenological Research, 86(1), 1–40.

Madison, B. J. C. (2014). Epistemological disjunctivism and the new evil demon. Acta Analytica, 29(1), 61–70.

Madison, B. J. C. (2017). On justification and excuses. Synthese, 1(12), 1–12.

Majors, B., & Sawyer, S. (2005). The epistemological argument for content externalism. Philosophical Perspectives, 19, 257–280.

McCain, K. (2016). Evidentialism and epistemic justification. London: Routledge.

McDowell, J. (1982). Criteria, defeasibility and knowledge. Proceedings of the British Academy, 68, 455–479.

McGlynn, A. (2014). Knowledge first?: Basingstoke: Palgrave Macmillan.

McNaughton, D., & Rawling, P. (2018). Motivating reasons and normative reasons. In D. Star (Ed.), Oxford handbook of reasons and normativity. Oxford: Oxford University Press.
Plantinga, A. (1993). *Warrant: The current debate*. Oxford: Oxford University Press.

Pollock, J. L., & Cruz, J. (1999). *Contemporary theories of knowledge*. Lanham: Rowman & Littlefield.

Pritchard, D. (2012). *Epistemological disjunctivism*. Oxford: Oxford University Press.

Pryor, J. (2001). Highlights of recent epistemology. *British Journal for the Philosophy of Science*, 52(1), 95–124.

Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.

Smithies, D. (2012). Mentalism and epistemic transparency. *Australasian Journal of Philosophy*, 90(4), 723–742.

Smithies, D. (2014). The phenomenal basis of epistemic justification. In M. Sprevak & J. Kallestrup (Eds.), *New waves in philosophy of mind* (pp. 98–124). Basingstoke: Palgrave Macmillan.

Smithies, D. (forthcoming). *The epistemic role of consciousness*. Oxford University Press. [Downloaded: 10.08.2017.]

Snowdon, P. (1980–1981). Perception, vision and causation. In *Proceedings of the aristotelian society* (Vol. 81, pp. 175–192).

Smithies, D. (2014). The phenomenal basis of epistemic justification. In M. Sprevak & J. Kallestrup (Eds.), *New waves in philosophy of mind* (pp. 98–124). Basingstoke: Palgrave Macmillan.

Smithies, D. (forthcoming). *The epistemic role of consciousness*. Oxford University Press. [Downloaded: 10.08.2017.]

Sutton, J. (2007). *Without justification*. Cambridge, MA: MIT Press.

Turri, J. (2009). The ontology of epistemic reasons. *Nouss*, 43(3), 490–512.

Turri, J. (2010). On the relationship between propositional and doxastic justification. *Philosophy and Phenomenological Research*, 80(2), 312–326.

Turri, J. (2015a). Evidence of factive norms of belief and decision. *Synthese*, 192(12), 4009–4030.

Turri, J. (2015b). The radicalism of truth-insensitive epistemology: Truth’s profound effect on the evaluation of belief. *Philosophy and Phenomenological Research*, 93(2), 348–367.

Warfield, T. (2005). Knowledge from falsehood. *Philosophical Perspectives*, 19, 405–416.

Wedgwood, R. (2002). Internalism explained. *Philosophy and Phenomenological Research*, 65(2), 349–369.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Williamson, T. (2007). On being justified in one’s head. In M. Timmons, J. Greco, & A. Mele (Eds.), *Rationality and the good: Critical essays on the ethics and epistemology of Robert Audi* (pp. 106–122). Oxford: Oxford University Press.

Williamson, T. (2016). Justifications, excuses, and sceptical scenarios. In J. Dutant & D. Dohrn (Eds.), *The new evil demon*. Oxford: Oxford University Press.