

## 1. Introduction

Consider:

**(JFB)** False beliefs can be justified<sup>1</sup>

JFB appears to draw support from two kinds of case. In what I'll call 'type-1' cases, one's experiences are *globally* misleading: given an experience with content  $p$ , it's the case that not- $p$ , because nothing is as it seems. In what I'll call 'type-2' cases, one's experiences are locally misleading: given an experience with content  $p$ , it's the case that not- $p$ , because things aren't as they seem in the situation, though they are sometimes as they seem. Here's a token of each:

Type-1: A experiences having hands. So A believes that she has hands. Unfortunately, A's experiences are simulations fed to her by a supercomputer: A is a *brain-in-a-vat* (*BIV*) (see, e.g., Lehrer & Cohen 1983: 192; Cohen 1984: 281; Sosa 1988: 164).

Type-2: B is at the garden center. She has the visual experience of a potted plant,  $p$ , in the 'real plants' aisle. So B believes that  $p$  is a real plant. Unfortunately,  $p$  is not a real plant: it's a plastic counterfeit (see, e.g., Lasonen-Aarnio 2021a: 107; Comesaña 2020: 1; Greco 2021: 519; Hughes 2017: 159).

A and B have false beliefs. So they are bad, in a way. But they are good in a way too. B's belief is hardly negligent: she's in the 'real plants' aisle, and every plant but  $p$  is real. A's belief isn't just passable, but *exemplary*: by taking it for granted that her senses put her in touch with the facts when she has, and could have, no indication to the contrary, A's case serves as an example of good perceptual believing. Type-1 and type-2 cases are collectively referred to as *bad cases*. Bad cases are to be contrasted with *good cases*, where one's experiences aren't misleading: in good cases, the belief turns out to be true. There's something to be said for beliefs in bad cases. But what should we say about them? Are they justified?

Here's the plan for the paper. §2 states the standard argument for JFB. §3 rehearses Timothy Williamson's (forthcoming A) objection to the standard argument.<sup>2</sup> §4 argues that this isn't, in fact, an objection to the standard argument: Williamson assumes that the argument cites compliance with one norm, but it cites compliance with a different norm. It isn't my goal to argue that JFB is true. Instead, I want to get clear on the most charitable way of representing the standard argument in Williamson's framework. The reward for doing so is a new perspective on the intuition that false beliefs can be justified: if the circumstances are right, a *false* belief can manifest a disposition to know, or to have only true beliefs.

---

<sup>1</sup> In this paper, I'll be concerned with *doxastic* justification; justification which applies, or fails to apply, to the beliefs of a subject in a situation.

<sup>2</sup> References to 'Williamson' are to Williamson (forthcoming A).

## 2. *The Standard Argument*

We have seen that beliefs in bad cases are somehow good. But if we're going to hang our case for JFB on bad cases, then we need some reason to think that they are not just *good*, but *justified*. As a rough starting point, there's the sense that, in forming or retaining her false belief, '[a BIV] is following the very cognitive instincts [they] ought to have' (Williamson: 19-20; see, also, Williamson 2017: 271). The idea applies to type-2 cases as well: in believing that *p* is a real plant, B is following the cognitive instincts she ought to have (compare: B believes that *p* is a fake plant, despite the fact that she sees that she's in the 'real plants' aisle and has no evidence that this is misleading).

According to the standard argument for JFB (henceforth 'SA'), if a belief is formed in (near enough) the *best* possible way, then it's justified:<sup>3</sup>

(SA)

(1) If a belief is formed<sup>4</sup> in (near enough) the best feasible<sup>5</sup> way,<sup>6</sup> then it is justified.

(2) Beliefs in bad cases are formed in (near enough) the best feasible ways.

(3) Beliefs in bad cases are justified [by (1) & (2)].

(JFB) There can be justified false beliefs [by (3)].

The argument is schematic, and deliberately so: SA is meant to subsume more specific cases for JFB. Proponents of SA have to say what they mean by 'feasible' and '(near enough) best.' First: 'feasible.' With Lasonen-Aarnio (2021a: 101-103; 2021b: 461-462; forthcoming B: 12), we could understand feasibility as a constraint flowing from our cognitive limitations. For example: a lifeguard on duty should pay attention, but summer reading is more pleasant. It would be nice for lifeguards, therefore, if they could be disposed to pay attention *only when* someone is about to drown. Unfortunately, insofar as lifeguards are human beings, they have no way to tell whether someone is going to drown until it's too late to stop it from happening. Paying attention only when someone is in imminent danger is a way of preventing people from drowning that is not *feasible* in Lasonen-Aarnio's sense.

This is one way to go; but it is not the only way. Beliefs in type-2 cases are not formed in the best feasible ways on this account. After all, it wasn't human cognitive limitations that prevented B from forming the belief that *p* is a fake plant. It would have been better if she had looked harder; and it would have been feasible for her to do so on Lasonen-Aarnio's picture. In order to get the

---

<sup>3</sup> See, *inter alia*, Cohen (1984: 283), Wedgwood (2002: 349); Bird (2007: 84), Simion et al. (2016: 282-283) and Ichikawa (2017: 115).

<sup>4</sup> To streamline discussion, I'll focus on belief-formation. But we could run a similar argument for belief-retention.

<sup>5</sup> I owe the terminology of 'feasible ways' to Maria Lasonen-Aarnio. See her (ms).

<sup>6</sup> For my purposes I'm happy to use 'way' pretty loosely. Because I take ways to be *dispositions* (see 4.1), I can avoid worries about how to individuate methods and processes (cf. Bird 2007: 88).

verdict that B's belief is justified, then, we'll have to show why, though not strictly speaking *best*, B's way is *good enough*. Another way to go would be to beef up the feasibility requirement. This would allow us to say that it is *not* feasible for B to take a long, hard look at the plastic plant. A suitably beefed-up notion of feasibility will incorporate considerations of expediency, or other 'non-epistemic' considerations, e.g., stakes. I won't argue for either alternative here.

That brings us to the issue of what makes a way good enough, or '(near enough) best.' Proponents of JFB are divided: some hold that 'external' goods like truth or knowledge are entirely irrelevant to the question whether a way is good enough; others hold that such considerations *are* relevant, but that a way needn't issue true beliefs or knowledge *all of the time* to be good enough. For example: some think that beliefs formed in (near enough) the best feasible ways fit the evidence (Conee & Feldman 2004); others, that they are formed in a way that is *reliable* at normal worlds; i.e., worlds where things are generally as they seem (Goldman 1986: 107-109); others, that they result from the *normal functioning* of the belief-forming process (Graham 2012). Each of these views offers a specification of '(near enough) best' which supports the verdict that beliefs in bad cases are justified: according to each, whether a belief is justified is independent of its truth in the subject's situation, and dependent on conditions satisfied in her situation.

What, if anything, is wrong with SA? According to Williamson, (1) is false. The next section begins with a primer on norms of belief, then turns to the objection.

### 3. Williamson's objection

#### 3.1 Norms of belief

Williamson assumes that justification is about demonstrating compliance with *norms*, where a norm is any rule issuing a *prima facie* 'ought' (Williamson: 4). Williamson (7–9) distinguishes three kinds of norm. The *primary* norm of belief is of the form:

(S)<sup>7</sup> one ought to (form a belief only if that belief satisfies C)

C is an objective, or 'truth related,' condition; for example, truth or knowledge (Williamson: 12): an instance of S is 'one ought to (form a belief that *p* only if *p* is true)' (the *truth* norm). Once you're in a situation, S tells you what (*prima facie*) you ought (not) to believe, or continue to believe, in that situation. When your belief satisfies C, you satisfy S's compliance condition (or, as I'll sometimes say, satisfy S). Suppose, for illustration, that C requires truth. If it's raining on Walden Pond, and you in fact form a belief about whether it is, then S says that you ought to believe that it's raining on Walden Pond.

From the standpoint of S, you should also be disposed to believe whatever S sanctions. So says the *secondary* norm of belief:

---

<sup>7</sup> Williamson labels the norm 'J,' for 'justifying.' But this prejudges the question whether the primary norm is justifying. So, I've replaced 'J' with 'S,' for 'successful.'

(DS) one ought to (be disposed to satisfy S)

DS says: 'be an S-er' (i.e. someone with that disposition). It is *derivative* of S in the following sense: you ought to be disposed to only form true beliefs (i.e., be so disposed that *if* you form a belief, then it's true) *because* you ought to only form a belief if true. As Williamson (8) stresses, we can be S-ers without *always* complying with the primary norm: we're fallible; we don't always believe what our ideal counterparts would believe. But it *is* necessary that our beliefs *normally* satisfy S. More could be said about this notion of 'normality', but Williamson draws the line here.

Finally, you should believe what an S-er would believe in your situation. So says the *tertiary* norm of belief:

(ODS) one ought to (believe what an S-er would believe in one's situation)

ODS is *occurrent* because its satisfaction is determined by what the subject *does* in her situation (see Williamson: 28). It is derivative of S in the following sense: you ought to do what someone disposed to only form true beliefs would do *because* you ought to only form a belief if true. If it's raining on Walden Pond, then you ought to believe, or continue to believe, what someone who tends to have only true beliefs would believe in your situation, if they were to consider the matter (that it's raining on Walden Pond).

(1), recall, says that one's belief is justified if formed in (near enough) the best feasible way. Which norm is at issue here? For Williamson (20), it's ODS: (1) says, in effect, that one's belief is justified if it's what an S-er would believe in one's situation. But this is false. Why?

### 3.2 Scrambler

First, Williamson (15) asks us to assume a minimal interpretation of S on which it rules out belief in contradictions (call it  $S_{\min.}$ ). This is a condition that he and his opponents will take to be the core of any plausible primary norm of belief (setting dialetheia aside). Consider:

**Scrambler:** Connie is disposed to comply with  $S_{\min.}$ . Exposed to a scrambler – a device which sporadically alters the contents of her working memory, temporarily taking her capacity to detect contradictions offline – she comes to believe that 179 *is* and *is not* prime (Williamson: 15-17; Williamson 2017: 268- 269).

Williamson reasons as follows: Connie violates  $S_{\min.}$  (she believes a contradiction). She satisfies  $DS_{\min.}$  by stipulation. Since Connie satisfies  $DS_{\min.}$ , whatever she believes will be what someone with the disposition to comply with  $S_{\min.}$  would believe. So she complies with  $ODS_{\min.}$  in forming her belief. We have it that Connie's belief violates  $S_{\min.}$  and satisfies  $ODS_{\min.}$ . That means the compliance conditions for S and ODS will diverge if S incorporates a consistency requirement. And we think that it will. But our verdict about Scrambler is that Connie's belief isn't justified. So compliance with ODS doesn't *justify*: if it did, then Connie's belief would be justified.

Why does this result threaten SA? Recall that, for Williamson, (1) says that if you believe what an S-er would believe in your situation, then you're justified. But we just saw that compliance with ODS doesn't justify. So SA doesn't establish JFB. That's Williamson's argument.

#### 4. *What the tertiary norm of belief is*

##### 4.1 *SA does not appeal to ODS*

Here's the problem with Williamson's argument: the norm to which SA appeals is *not* ODS. Recall the major premiss:

(1) If a belief is formed in (near enough) the best feasible way, then it is justified.

(1) does not say that ODS-compliance is justifying. If ODS were at issue, then the premiss would read:

(1\*) If a belief is formed in the way that an S-er would form her belief, then it is justified.

But (1\*) is implausible: why should the fact that one's way matches an S-er's way justify the resulting belief if one's situation is such that *even an S-er's* way would be defective by S's lights?<sup>8</sup> The answer is that it shouldn't: *no one* (I suppose not even Connie) thinks that Connie's belief in Scrambler is justified. That strongly suggests that ODS isn't the norm at issue. Instead, SA appeals to a norm of the following kind:

(ODS<sup>+</sup>) one ought to (*manifest a disposition* to comply with S)<sup>9</sup>

Manifestation is a familiar notion. A glass manifests fragility when shattered by a soprano; a match manifests inflammability when it lights; and a castle manifests vulnerability when reduced to rubble by canon fire (Williamson: 22; 2017: 272). An object can have a disposition without manifesting it. A fragile glass may never be heated, and so never crack; never be sung to, and so never shatter and so on. Moreover, the glass may be wrapped in protective packaging, or guarded by a sorcerer capable of producing shockwaves which prevent it from shattering upon being struck (Bird 1998: 223). In such cases, even being struck, e.g., does not cause fragility to manifest: the disposition is *masked*.<sup>10</sup> From this point, my argument will be simple: (1) says that dispositional manifestation justifies; Connie's S-compliant disposition is masked in Scrambler; so Scrambler isn't a counterexample to (1).

First, why may (1) be glossed as 'ODS<sup>+</sup> justifies' but not as 'ODS justifies'?<sup>11</sup> (1) says that being formed in (near enough) the best feasible way is justifying for belief. But even S-ers can fail to form

---

<sup>8</sup> See Lasonen-Aarnio (forthcoming A: 7; 2021b: 462) for essentially the same complaint.

<sup>9</sup> This reading is assumed by Hughes (2019: 64-65) and discussed, briefly, by Boulton (2017: 4076) and Brown (2018: 90).

<sup>10</sup> On masking, see McKittrick (2018: 26), Fara (2005: 50), Bird (1998: 228) and Johnston (1992: 233).

<sup>11</sup> Williamson (p.c.) would add a fourth, hybrid norm, according to which one should satisfy the secondary and tertiary norms:

beliefs in ways that are good by the lights of S (e.g. when they're tired, hungry, or having their brains scrambled). So it is, strictly speaking, irrelevant what an S-er would believe. What is relevant is what manifesting the best dispositions would come to; which way of forming a belief would be best by the lights of S given what's feasible in the situation.<sup>12</sup>

Recall that the best feasible ways are not going to deliver truth or knowledge all of the time at every world: even our best ways of acquiring knowledge – believe your eyes, or the experts – sometimes lead to error. So a disposition that manifests as forming beliefs in (near enough) the best feasible ways will sometimes deliver a belief that violates S. In what sense, then, is the disposition in question 'a disposition to *comply* with S'? In type-2 cases, where one's evidence is only *locally* misleading, it's plain that manifesting this disposition implies S-compliance in *normal* cases, understood as cases where things *are as they seem* given non-deficient evidence: plastic plants in the 'real plants' aisle are an abnormal occurrence in this sense; so B's belief-forming process will deliver true or knowledgeable beliefs more often than not. Things are different in type-1 cases, where the beliefs manifest a disposition that, in the *subject's* world, *always* delivers false and ignorant beliefs. But, on any reasonable understanding of feasibility, it isn't feasible for subjects in type-1 cases to have true or knowledgeable beliefs. So it doesn't count against a disposition's being an *S-compliant* disposition that it fails to manifest as S-compliance at such worlds. What matters is that the disposition usually manifests as S-compliance at *normal worlds*; i.e., worlds relevantly like the world as we take it to be.<sup>13</sup> Williamson (14) seems to agree: even BIVs comply with DS and ODS (for all that he says, they comply with ODS<sup>+</sup> too<sup>14</sup>). Just as being a promise-keeper doesn't imply that you normally keep your promises at every world, so being an S-er doesn't imply that you normally form true or knowledgeable beliefs at every world (some, such as A's, are too disabling).

This is the sense in which the disposition to form beliefs in the best feasible ways is a disposition to comply with S: at a normal world, if you manifest that disposition, you'll mostly form S-compliant beliefs. And the existence of abnormal worlds doesn't affect the goodness of the disposition from the standpoint of S. In exactly the same way, the existence of possible worlds where a demon prevents anyone from keeping their promises doesn't threaten the goodness of a promise-keeping disposition from the standpoint of promise-keeping.

---

(MDS) one ought to (believe (what an S-er would believe) *because* one is an S-er)

MDS says that one should manifest *one's* disposition to comply with S. Since, unlike ODS and ODS<sup>+</sup>, MDS requires one to *be* an S-er, MDS is not occurrent. Why do I pursue J = ODS<sup>+</sup> rather than J = MDS? J = MDS turns out to be *too stingy*: even a belief formed in an immaculate way may fail to be justified if the subject happens not to be an S-er; i.e., if their immaculately-formed belief is aberrant for them. This is implausible. The justifying norm will require dispositional *manifestation*, but it will also have to be *occurrent*. Only ODS<sup>+</sup> satisfies both desiderata.

<sup>12</sup> I'm not claiming that all proponents of the standard argument would agree that (1) cites compliance with ODS<sup>+</sup>: indeed, many do not appeal to dispositions in the first place. Rather, the claim is that *within* Williamson's framework, this is the kind of norm that best respects (1).

<sup>13</sup> See Lasonen-Aarnio (2010: 13; forthcoming A: 16; forthcoming B: 11-12).

<sup>14</sup> Indeed, Williamson (forthcoming B: 7) grants that the victim of a type-1 case 'exercises knowledge-conducive cognitive dispositions.'

#### 4.2 ODS<sup>+</sup> is the tertiary norm of belief

So far, I have argued for a more charitable reading of SA on which it cites compliance with ODS<sup>+</sup>. ODS and ODS<sup>+</sup> recommend the same beliefs in cases where an S-er would manifest her S-compliant disposition. For example, if B is an S-er, then her belief satisfies ODS<sup>+</sup> and ODS: it is formed in a good way in circumstances in which that's what we'd expect from her. So, in this kind of case, there's no reason to prefer ODS<sup>+</sup> to ODS. If we only consider such cases, then we'll end up running the norms together because we'll never consider the possibility of masking. When an S-compliant disposition is masked, however, the resulting belief is not plausibly justified.

An interpreter of Chopin may ask herself: 'what would Horowitz do?' And this may be a productive exercise; her interpretation may be the better for it. But what if Horowitz were fingerless or crippled by stage fright? What then? What our pianist must be doing, for the question to be productive, is *bracketing* these possibilities, so that Horowitz's virtuosity is *manifest* to her. Returning to belief, my suggestion is that when we ask ourselves 'which attitude would an S-er take?' we care about what an S-er would do only insofar as their S-compliant disposition is manifest to us. In other words, the only reason to care about what an S-er would do is that an S-er is someone who normally manifests an S-compliant disposition. So we have the following order of priority: you should do what an S-er would do *because* you should manifest an S-compliant disposition. Now, this is just to say that ODS is *derivative* of ODS<sup>+</sup>. Recall that the tertiary norm of belief is the occurrent norm derivative of S. Unless there is some *even more* fundamental occurrent norm derivative of S,<sup>15</sup> then, ODS<sup>+</sup> is the tertiary norm of belief.<sup>16</sup>

#### 4.3 Williamson's objection is ineffective

I have argued that (1) requires dispositional manifestation. Is Connie's S-compliant disposition manifest in Scrambler? It is not: the scrambler blocks its manifestation, the way carbon dioxide prevents a match from lighting. Perhaps Connie's calculations show great mathematical sophistication; perhaps she exhibits plenty of understanding of what it takes to be prime and what it is to be the number 179. Even so, by believing what she does, she exhibits insensitivity to the conspicuous fact that there are no true contradictions. She thereby manifests a disposition to have lots of false mathematical beliefs. In short: if Connie complies with ODS<sup>+</sup>, then that shows that ODS<sup>+</sup> isn't justifying. But Connie doesn't comply with ODS<sup>+</sup>. So we'll need some other reason to think that ODS<sup>+</sup> isn't justifying.

I have argued that (1) cites compliance with the norm ODS<sup>+</sup>. Since ODS<sup>+</sup> is violated in Scrambler, Scrambler is not a counterexample to (1). For all I have said, beliefs satisfying ODS<sup>+</sup> might be

---

<sup>15</sup> While I don't see how there could be such a norm, I can't go into the matter here.

<sup>16</sup> As an anonymous referee points out, I haven't argued that ODS<sup>+</sup> is a *genuine* norm. Of course, anyone who likes SA *and* thinks that justifications demonstrate compliance with all of the norms will want to say that it is. But I'm not arguing that here. Instead, I'm arguing for the conditional claim that *if* you should comply with ODS, that's *because* you should comply with ODS<sup>+</sup>. I'm taking a stand on their relative, rather than their absolute, status.

merely reasonable.<sup>17</sup> Against this, it could be argued that beliefs satisfying ODS<sup>+</sup> are justified.<sup>18</sup> A natural next step in the defense of JFB would be to make this argument.<sup>19</sup>

*School of Philosophical, Anthropological and Film Studies,  
The University of St. Andrews  
United Kingdom  
[jd259@st-andrews.ac.uk](mailto:jd259@st-andrews.ac.uk)*

## References

- Bird, A. 2007. Justified Judging. *Philosophy and Phenomenological Research* 74: 81–110.
- Bird, A. 1998. Dispositions and Antidotes. *The Philosophical Quarterly* 48: 227–34
- Boult, C. 2017. Epistemic normativity and the justification-excuse distinction. *Synthese* 194: 4065–81.
- Brown, J. 2018. *Fallibilism: Evidence and Knowledge*. Oxford: Oxford University Press.
- Cohen, S. 1984. Justification and Truth. *Philosophical Studies* 46: 279–95.
- Comesaña, J. 2020. *Being Rational and Being Right*. Oxford: Oxford University Press.
- Conee, E. and Feldman, R. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Clarendon Press.
- Fara, M. 2005. Dispositions and Habituals. *Noûs* 39: 43–82.
- Goldman, A. 1986. *Epistemology and Cognition*. Cambridge: Harvard University Press.
- Goodman, J. 2013. Inexact Knowledge without Improbable Knowing. *Inquiry* 56: 30–53.
- Graham, P. 2012. Epistemic Entitlement. *Noûs* 46: 449–82.
- Greco, D. 2021. Justifications and Excuses in Epistemology. *Noûs* 55: 517–37.
- Hughes, N. 2019. Uniqueness, Rationality, and the Norm of Belief. *Erkenntnis* 84: 57–75.
- Hughes, N. 2017. No Excuses: Against the Knowledge Norm of Belief. *Thought* 6: 157–66.
- Ichikawa, J. 2017. *Contextualising Knowledge: Epistemology and Semantics*. Oxford: Oxford University Press.
- Johnston, M. 1992. How to Speak of the Colors. *Philosophical Studies* 68: 221–63.
- Lasonen-Aarnio, M. ms. *The Good, The Bad, and the Feasible: Knowledge and Reasonable Belief*.
- Lasonen-Aarnio, M. forthcoming A. Competent Failure and Victims of Deceit. In *The New Evil Demon*, eds. J. Dutant and F. Dorsch. Oxford: Oxford University Press.
- Lasonen-Aarnio, M. forthcoming B. Perspectives and Good Dispositions. *Philosophy and Phenomenological Research*.
- Lasonen-Aarnio, M. 2021a. Dispositional Evaluations and Defeat. In *Reasons, Justification, and Defeat*, eds. J. Brown and M. Simion, 93–115. Oxford: Oxford University Press.
- Lasonen-Aarnio, M. 2021b. Coherence as Competence. *Episteme* 18: 453–76.
- Lasonen-Aarnio, M. 2014. Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88: 314–45.
- Lasonen-Aarnio, M. 2010. Unreasonable Knowledge. *Philosophical Perspectives* 24: 1–21.

---

<sup>17</sup> For this view, see Lasonen-Aarnio (2010: 13; 2014: 343; 2021a: 99).

<sup>18</sup> Goodman (2013: 50–52) expresses sympathy for this claim.

<sup>19</sup> I'm grateful to Jessica Brown, Patrick Greenough, Evie Moss, Brian Rabern, and three anonymous referees for written feedback on earlier drafts.



- Lehrer, K. and Cohen, S. 1983. Justification, Truth, and Coherence. *Synthese* 55: 191–207.
- McKittrick, J. 2018. *Dispositional Pluralism*. Oxford: Oxford University Press.
- Simion, M., Kelp, C. and Ghijsen, H. 2016. Norms of Belief. *Philosophical Issues* 26: 374–92.
- Sosa, E. 1988. Beyond Scepticism, to the Best of our Knowledge. *Mind* 97: 153–88.
- Wedgwood, R. 2002. Internalism Explained. *Philosophy and Phenomenological Research* 65: 349–69.
- Williamson, T. forthcoming A. Justifications, Excuses, and Sceptical Scenarios. In *The New Evil Demon*, eds. J. Dutant and F. Dorsch. Oxford: Oxford University Press.
- Williamson, T. forthcoming B. Epistemological Ambivalence. In *Epistemic Dilemmas*, ed. N. Hughes. Oxford: Oxford University Press.
- Williamson, T. 2017. Ambiguous Rationality. *Episteme* 14: 263–74.