

Boghossian, Müller-Lyer, the Parrot, and the Nazi
 (penultimate draft of paper to appear in Luis Oliveira (ed.), *Externalism about Knowledge*.
 Oxford: Oxford University Press)

Timothy Williamson

Distinctions between internalism and externalism draw inspiration from the idea that the mind has an *inside* and an *outside*, literally or metaphorically. Internalism about a feature is the view that it is constitutively determined by what is inside the mind; externalism about it is the denial of such internalism. The distinction between inner and outer may be drawn either mentalistically, typically in terms of how things seem to the subject, or materialistically, typically in terms of what goes on in the subject's brain. One asymmetry between internalism and externalism is that the distinction between inner and outer is normally central to internalism, but marginal to externalism. For the claim of determination by the inner is at the heart of the internalist's positive account, whereas the denial of determination by the inner is a mere by-product of the externalist's positive account, which is typically articulated otherwise: for example, in terms of causation or knowledge. Since internalists *need* such a distinction much more than externalists do, it is only fair to leave internalists to draw it as they see fit.

In practice, of course, disputes between internalists and externalists are often conducted in more specific terms, with no officially agreed inner-outer distinction. But some such contrast is usually there, in the background, motivating the internalist. That applies in particular to the exchanges between Paul Boghossian and me on epistemic justification (Boghossian and Williamson, 2020). Boghossian's arguments manifest an internalist outlook, with a mentalistic rather than materialistic inner-outer distinction in the background: consciously accessible seemings to the subjects are central to his account of epistemic justification. For perceptual beliefs, the relevant seemings are sensory. For moral beliefs, they are *intellectual*. Boghossian draws a strong analogy between the role of sensory seemings in justifying perceptual beliefs and the role of intellectual seemings in justifying many non-perceptual beliefs, including moral beliefs.

One obvious challenge to such internalist accounts of what it takes for beliefs to be justified is that they threaten to justify the most depraved bigots' nastiest beliefs, especially their moral beliefs, since they cohere so well with the bigots' intellectual seemings. A similar challenge arises to analogous internalist accounts of what it takes for *actions* to be justified. I raised such concerns in the book, but did not press them hard, because Boghossian left his positive account of epistemic justification in a provisional and rather undeveloped form. His paper 'Intuition and A Priori Justification' (Boghossian, 2021) slightly modifies his account, but also responds to my criticisms, both of the justificatory role he assigns to seemings in general and of its implications for the justification of bigoted beliefs in particular. In the present paper, I use his response to press my objections further.

Section 1 explains some general problems for Boghossian's account of the justification of perceptual beliefs. In the paper under discussion, he makes extensive use of the Müller-Lyer illusion to support his view of perception; section 2 shows why the illusion does not help him. Section 3 explains some general problems for Boghossian's account of the justification of moral beliefs, analogous to the problems in section 1 for perceptual beliefs. Section 4 applies the preceding discussion to the (alleged) justification of bigoted beliefs, and then raises problems for Boghossian's attempt to stop justifications of bigots' beliefs from spreading into justifications of bigots' actions.

1. *Sensory seemings*

Here is Boghossian's initial account of sensory seemings (all quotations from Boghossian (2021) unless otherwise specified; all italics in them Boghossian's):

As I understand it, a sensory seeming is a state, with both propositional content and a sensory phenomenology, which directly *presents* a proposition as true. 'Direct' is strictly redundant given 'presents': to present p as true is to display it as true in a direct, non-inferential way.

He adds:

Such a sensory seeming is not itself a judgment, belief or a disposition to believe; it is pre-doxastic. However, because such a sensory seeming presents a proposition as true, it typically *results* in an inclination to believe that proposition (unless that inclination is checked by background knowledge).

For vision, Boghossian calls sensory seemings 'visual impressions'. He states their epistemic role thus:

visual impressions are often the *consciously accessible reasons for which* we not only believe something, but are disposed to believe it. Furthermore, these visual impressions are the justifiers for our believing their contents, when we do so.

What evidence is there that we have sensory seemings, as Boghossian understands them? Imagine a normal case where you see that a car is coming towards you. On Boghossian's model, your perceptual belief that a car is coming towards you is justified by your visual impression that a car is coming towards you. That visual impression is distinct from your seeing that a car is coming towards you, for being in the latter state entails that a car *is* coming towards you (seeing-that is factive), whereas for Boghossian you can have the visual impression even when it is illusory. Your visual impression *that a car is coming towards you* must also be distinguished from your total visual experience at the time, for the latter subsumes many other visual impressions too, each with its own propositional content, for example, your visual impression *that the car is red*. Still, we can agree that it does visually appear to you *that a car is coming towards you*; the state of being so appeared to is non-factive and has the appropriate propositional content. Why cannot that state serve as the visual impression in Boghossian's sense?

The problem is that Boghossian requires visual impressions to be *pre-doxastic*: they must not consist in beliefs, or even dispositions to believe (though they may cause them). For, he reasonably fears, if visual impressions were doxastic, they would be circular as justifiers for the corresponding beliefs with the very same contents. In justifying your belief that a car is coming towards you, a doxastic state with the propositional content that a car is coming towards you will in effect be justifying itself. So the question becomes: what evidence is there that the state of its visually appearing to you that a car is coming towards you is pre-doxastic, that it can visually appear to you that a car is coming towards you even though you have no disposition to believe that a car is coming towards you, let alone an actual belief to that effect? Of course, you might resist forming the belief because you suspect that some illusion is at work, but a resisted disposition is still a disposition.

When I look for Boghossian-style pre-doxastic visual impressions in myself, I cannot find them. It sometimes visually appears to me that a car is coming towards me, but those are simply occasions when vision disposes me to believe that a car is coming towards me, whether or not I actually form the belief. The screw is tightened by Boghossian's insistence on the conscious accessibility of visual impressions (see the last quoted passage above). If no pre-doxastic visual impressions are consciously accessible to me, I have no Boghossian-style visual impressions. Of course, I might be so blinkered by philosophical prejudice that I cannot recognize what is right in front of my mind's eye for what it is, but the onus is on Boghossian to show that something meets his list of requirements: a pre-doxastic, consciously accessible visual impression which presents the relevant proposition as true.¹

Boghossian's strategy starts from the (correct) observation that when we form a visual belief, it normally does not feel like 'taking a stab in the dark' or 'guessing'. His alternative is that we form the belief for a (good) consciously accessible reason, and his candidate reason is that we have the visual impression (in his sense). Thus you form the belief that a car is coming towards you for the reason that you have the visual impression that a car is coming towards you. The visual impression must be pre-doxastic, otherwise the reason would be a blatantly bad one, because blatantly circular. The visual impression must also be consciously accessible, for if it were hidden, forming the belief should still feel like taking a stab in the dark or guessing, which it does not.

Phenomenologically, Boghossian's story does not ring true. He postulates a psychological process in which having the visual impression causes one to form a belief with the same content. Call that process *the internalist move*. Boghossian needs the move to be made even when one confidently and unhesitatingly forms perceptual beliefs, otherwise such beliefs would not be based on reasons in the way he requires. I for one am quite unaware of making such internalist moves when I confidently and unhesitatingly form a stream of perceptual beliefs. Even in retrospect, I am quite unaware of making such moves. In that respect, I doubt that my phenomenology is radically different from the reader's, or even from Boghossian's. Suppose, for the sake of argument, that nevertheless we are in fact almost always making the internalist moves, but so quickly that we almost never notice doing so. Then, by Boghossian's lights, we should expect forming confident, unhesitating perceptual beliefs *usually* to feel like taking a stab in the dark or guessing, since we fail to notice the reasons—the pre-doxastic visual impressions—which are supposed to differentiate normal perceptual belief formation from such activities. But, as Boghossian himself insists, normal

perceptual belief formation does *not* feel like taking a stab in the dark or guessing. Thus Boghossian's internalism flunks the test he sets externalism.

Evolutionarily, Boghossian's story makes as little sense as it does phenomenologically. Consciously accessible processing is notoriously slow; efficiency requires most cognitive processing to be automatic and unconscious. Boghossian postulates a cognitive architecture on which the most ordinary perceptual belief formation requires a consciously accessible internalist move, for no discernible practical benefit. That would create a gratuitous cognitive bottleneck in processing the normal stream of perceptual beliefs. One would expect creatures with cognitive systems built on the lines proposed by Boghossian, before they could believe their eyes, to be eaten by predators.

Of course, *sometimes* we really do hesitate, unsure what perceptual belief to form, considering the appearances. We strain our eyes, wondering what that animal in the distance is. In such exceptional cases, our conscious thought processes may really feel like taking a stab in the dark, or guessing. They are presumably cases where automatic, unconscious reasoning failed to resolve some significant issue, which had to be referred upstairs for more considered treatment. That makes practical, evolutionary sense; it also fits the phenomenology much better than Boghossian's indiscriminate account.

Naturally, internalists will ask whether externalists can do any better. Boghossian comments that they will be saddled 'with the burden of trying to give a plausible account of perceptual justification using only externalist resources – reliability, etc. – with all of that stratagem's known failings'. But a more direct line is open to knowledge-first externalists. In the good case, you *see that* a car is coming towards you, and in doing so *know that* a car is coming towards you (Williamson, 2000, pp. 33-41). The question of knowledge is prior to the question of justification. The case does not feel like taking a stab in the dark, or guessing, because it feels like *knowing*.² More specifically, it feels like knowing that something is so by seeing that it is so. Obviously, it may still feel that way even when something has gone wrong and you do not really see that something is so; it still does not feel like taking a stab in the dark, or guessing (see Williamson (2022) for a knowledge-first account of what passes for justification in such cases). Such an account fits the phenomenology much better than does a postulated cognitively accessible visual impression anticipating the propositional content of ordinary visual knowledge, forced on us by an impoverished conception of the theoretical alternatives.

Internalists sometimes take heart from the thought that, often, if you ask someone *why* she is disposed to believe that things are some way, she will respond that that's how they look (wording modelled on Boghossian's claim *Explanation*). The use of the words 'is disposed to believe' rather than the plain 'believes' already suggests that the subject is uncertain; her response has a strong flavour of backing off under pressure. Adverting to perceptual appearances is by no means a mandatory response to such questions. Often, it is more natural to give information about how one knows. Asked 'Why do you believe that he's unhappy?', someone more sensitive to body language may quite naturally respond 'I can *see* that he's unhappy'. Similarly, asked 'Why do you believe that the picture is crooked?', someone with better eyesight may properly respond 'I can *see* that it's crooked'. Appropriate responses depend on the purposes of the conversation, its common ground, the dialectical position, and so on. Requests for justification may also pressure respondents into confabulation, saying

whatever one can think of most likely to satisfy the questioner. As evidence for pre-existing structures of justification, such conversational patterns are frequently misleading.

2. *The Müller-Lyer illusion*

For visual impressions, Boghossian's prime exhibit is the Müller-Lyer illusion. Fairly uncontroversially, it provides a counterexample to the principle that whenever it visually appears to you that P, you believe that P. One line continues to look longer than the other even once you know independently that they are of equal length. What is controversial is whether the illusion also provides a counterexample to the weaker principle that whenever it visually appears to you that P, you are at least *disposed* to believe that P. For, even when you know independently that the lines are of equal length, your visual module could still give you the disposition to judge and believe that one line is longer than the other, which you inhibit because you know better. However, Boghossian claims that we could retain the visual impression that one line is longer than the other even after entirely losing the disposition to believe that one is longer than the other. He writes of the perceiver:

She may also, perhaps only after long experience with the diagram, lose any *inclination* to believe [that the lines are of unequal length]. [...] Still, the visual impression *persists*, unshakably presenting two lines as being of unequal length. Indeed, it's quite clear that the visual impression of unequal lines *never* goes away, no matter how much [she] may come to know, and grow used to, the fact that the lines are equal in length.

What Boghossian fails to provide is any evidence that the visual impression that the lines are of unequal length is *more* persistent than the disposition to believe that they are of unequal length. There is good evidence that the impression of unequal length can be reduced, in some cases almost to zero, by various cues and possibly by habituation, though the details are unclear (Mountjoy, 1958, Rudel and Teuber, 1963, Predebon, 1998, 2006, Millar and Al-Attar, 2002, Pepperberg, Vicinay, and Cavanagh, 2008), though Boghossian claims 'the effect itself never goes completely away and doesn't decrease as a mere result of familiarity with the illusion (see Predebon (2006))'. Evidently, these are questions for experimental psychology. He then says:

However, isn't it entirely clear that we *could* lose the disposition to believe [that the lines are of unequal length] as a result of intimate familiarity with the illusion? That would certainly appear to be metaphysically possible; indeed, it would appear to be even nomologically possible.

None of this amounts to a serious argument that it is nomologically or even metaphysically possible that one could have the visual impression that the lines are of unequal length with no disposition to believe that they are of unequal length. It is doubtless nomologically possible to look attentively at the Müller-Lyer diagram without forming a disposition to believe that the lines are of unequal length, but the experimental evidence does not remotely suggest that it is nomologically *impossible* to look attentively at the Müller-Lyer diagram without forming the visual impression that the lines are of unequal length, in

particular, without having the proposition that they are of unequal length presented as true. Of course, if one already assumes that the visual impression is prior to the disposition to believe, one will find it plausible that the former could occur without the latter; but that assumption is the very point Boghossian has to prove.

Perhaps aware of the weakness of these arguments, Boghossian adds another:

In addition to these considerations, a very serious blow to [the Constitutive View that where there is no disposition to believe that P, there is no visual impression that P] is dealt by the ample evidence that pigeons, parrots, and even insects are subject to the Müller-Lyer illusion, although presumably we should be reluctant to attribute dispositions to belief to these creatures.

For evidence, he cites Nakamura et al. (2006), Pepperberg et al. (2008), Srinivasan (1993), and Feng et al. (2017). Boghossian says nothing to support his presumption that ‘we should be reluctant to attribute dispositions to belief to these creatures’. After all, he seems to think that they have visual impressions in his sense, in which case propositions are presented as true to these creatures. But if propositions are presented as true to a creature, why should it be such a stretch to attribute to it belief in some of those propositions? On the other hand, if the creatures are *not* assumed to have visual impressions, what relevance are they supposed to have to the argument?

Boghossian’s low view of the creatures’ cognitive capacities is not entirely shared by the researchers whose work he cites. In particular, Pepperberg et al. (2008) concerns Alex, a Grey parrot with a repertoire of over 90 vocalizations (‘including labels for foods and locations’) to choose from in answering the investigators’ questions (ibid.: p. 772). The authors—Irene Pepperberg, Jennifer Vicinay, and Patrick Cavanagh—describe Alex as making ‘relative size judgments’, as having the ability ‘to understand the concept of relative size and absence of size difference’, and as being asked ‘to identify the bigger or smaller one [of two horizontal lines] or to state “none” if they were equally sized’ (ibid.: p. 767). Shown two differently coloured lines, Alex answered questions such as “What color bigger?” and “What color smaller?” from a repertoire of seven colour terms with an accuracy rate around 80%. There was some evidence that he lost the illusion after repeated exposure, possibly through habituation (ibid.: p. 777). Obviously, the investigators are in a much better position to judge Alex’s cognitive capacities than Boghossian and I are. Given what they say, it would be perverse to refuse to ascribe beliefs to Alex. For example, when Alex is presented with a blue line and a yellow line and asked “What color bigger?”, and he answers “Blue”, a natural explanation of his answer “Blue” would include the premise that Alex believes that the blue line is longer (“bigger”) than the yellow line. The investigators attribute judgments to Alex; there would be no good motivation for doing so while refusing to attribute beliefs to him.

One might be willing to attribute beliefs to parrots but unwilling to attribute them to insects. Indeed, in arguing that even insects experience visual illusions, Mandyam Srinivasan (1993) is mostly cautious in attributing mental states, given the comparatively small number of neurons in an insect brain. For example, he writes: ‘The finding that the Mueller-Lyer figures elicit similar patterns of fixation in humans and flies suggests (but does not prove) that flies experience the illusion just as we do’ (ibid.: p. 653) and ‘flies behave as though they experience the Mueller-Lyer illusion’ (ibid.: p. 654).³ Srinivasan’s results do not make it

plausible that the proposition that one line is longer than the other is presented as true to the fly (nor does he claim otherwise). The similarity he is tentatively suggesting between fly and human seems to be at a lower cognitive level than that of propositional content. Since he is not crediting flies with visual impressions in Boghossian's sense, their bearing on Boghossian's argument that a creature can have visual impressions without dispositions to believe is at best indirect. Perhaps flies have visual *proto-impressions*, but then they may also have dispositions to *proto-believe*.

There is a more general reason why experimental results about non-human animals are unlikely to help Boghossian's case. The evidence that such animals experience the Müller-Lyer illusion comes from their behaviour. They act as they would if they were acting on the premise that one line is longer than the other. But treating a proposition as a premise on which to act is the hallmark of belief! In effect, the evidence that the animals experience the illusion is that they act as they would if they believed that one line was longer than the other. That is no doubt putting it too crudely, but the point remains: the experiments are not designed to identify cases in which animals have the visual impression that one line is longer than the other without being disposed to believe that one is longer than the other. Such experiments are unlikely to help Boghossian.

All in all, Boghossian's considerations about the Müller-Lyer illusion provide no evidence of consciously accessible pre-doxastic visual impressions. Given their phenomenological and evolutionary implausibility, and the availability of alternative epistemologies with no need of them, we should reject the postulation of visual impressions as Boghossian understands them, and more generally the postulation of sensory seemings in his sense. That is also bad news for his postulation of *intellectual* seemings, which he motivates in part by analogy between intellectual and sensory seemings. If the cases are as similar as Boghossian suggests, the failure of one is evidence for the failure of the other.

Nevertheless, it would be lazy and imprudent to rest the whole case against intellectual seemings on the assumption that if Boghossian is wrong about sensory seemings, then he is also wrong about intellectual seemings. The second half of this paper assesses his postulation of intellectual seemings without relying on the case against his postulation of sensory seemings.

3. *Intellectual seemings*

Boghossian compares and contrasts intellectual seemings thus:

An intellectual seeming is similar to a sensory seeming in presenting a proposition as true; it is dissimilar to it in not necessarily having a sensory phenomenology.

An intellectual seeming is similar to a sensory seeming in being pre-judgmental; it is dissimilar to it in that a sensory seeming is capable only of presenting a proposition as true simpliciter, whereas an intellectual seeming is capable of presenting a proposition not only as true, but as false; and not only as true (false) but as *necessarily true (false)*.

Boghossian also uses the term ‘intuition’ for intellectual seemings, and ‘rational intuition’ for intellectual seemings which present their propositional content as necessarily true (false). He assigns intuitions a similar epistemic role to the one he assigns sensory seemings:

Intuitions are often the *consciously accessible reasons for which* we not only believe something, but are disposed to believe it. Furthermore, these intuitions are the justifiers for our believing their contents, when we do so.

Boghossian supports this account of intellectual seemings by consideration of an example:

(NoTorture) It is morally wrong to inflict pain upon a human being merely for one’s own amusement.

He argues:

If you are aware of no reason for being disposed to believe (NoTorture), you would be in the situation of a ‘blind intuiter’: just like the blind-sighter, you would experience the phenomenology of making the judgment as akin to taking a stab in the dark, of guessing.

However, that clearly isn’t the phenomenology of a typical subject judging NoTorture to be true. The judgment doesn’t feel like a stab in the dark; there appear to be consciously accessible reasons for making the judgment which explain your making it. What are those reasons?

Boghossian argues that there is often no cognitively accessible reason for judging NoTorture to be true other than an intellectual seeming of NoTorture itself (one whose propositional content is NoTorture). Hence, if that intellectual seeming were not itself a cognitively accessible reason for judging NoTorture to be true, there would be no cognitively accessible reason at all for judging NoTorture to be true. In that case, judging NoTorture to be true would feel like taking a stab in the dark. Since judging NoTorture to be true does not feel like taking a stab in the dark, Boghossian concludes that the intellectual seeming of NoTorture *is* a cognitively accessible reason for judging NoTorture to be true. As a corollary, the intellectual seeming of NoTorture must be pre-doxastic, otherwise it would not be a proper reason for judging NoTorture to be true, since the justification would be circular.

Just as in the perceptual case, Boghossian ignores the knowledge-first alternative. When one judges that it is morally wrong to inflict pain upon a human being merely for one’s own amusement, one has the feeling of *knowing* that it is morally wrong to inflict pain upon a human being merely for one’s own amusement. It does not feel like a metaphorical case of blind-sight, because it feels like a metaphorical case of sight (insight). One feels no immediate need for any further reason. As before, the question of knowledge is prior to the question of justification.

The simple knowledge-first account fits the phenomenology better than does the strained distinction between the consciously accessible pre-doxastic intellectual seeming and the consciously accessible judgment with the same content. Indeed, someone who convinces himself that it is morally wrong to inflict pain upon a human being merely for one’s own amusement by considering his own mental states (his intellectual seemings) is looking in the wrong place: it’s not about him, it’s about inflicting pain upon a human being merely for one’s own amusement.

Boghossian's phenomenological argument for a consciously accessible pre-doxastic intellectual seeming fails for NoTorture because it neglects the phenomenology of knowing. He applies the same argument structure to verdicts in Gettier thought experiments; it fails for the same reason (Nagel, 2012 is directly relevant to the feeling of knowing in epistemological matters).

A slightly more complex case is Boghossian's discussion of this principle:

(Phenomenal Sorites) If two things look exactly alike, then if one looks red, so does the other.

Boghossian argues for a pre-doxastic intellectual seeming of Phenomenal Sorites thus:

This principle not only looks true; it looks to be *epistemically analytically true* – true merely in virtue of being understood [footnote crediting Crispin Wright for the example]. However, we know that it is not true because, applied unrestrictedly, it leads to falsehood. Knowing this, we not only lose the belief that it is true, we lose even the disposition to believe that it is true. Even after we have abandoned any disposition to assent to it, though, it continues to look not only true, but epistemically analytically true.

Unfortunately, Boghossian's discussion of the example is vitiated by an overlooked crucial ambiguity in Phenomenal Sorites, more specifically, in the phrase 'look exactly alike' (compare Williamson 1990, p. 93 on the ambiguity of 'feel the same'). One reading concerns the sameness of appearance; the other concerns the appearance of sameness. On the sameness of appearance reading, 'look' is in the scope of 'exactly alike'; we might paraphrase ' x and y look exactly alike' as ' $\text{how } x \text{ looks} = \text{how } y \text{ looks}$ '. On the appearance of sameness reading, 'exactly alike' is in the scope of 'look'; we might paraphrase ' x and y look exactly alike' as ' $\text{it looks as though } (x \text{ and } y \text{ are exactly alike})$ '. The sameness of appearance reading makes Phenomenal Sorites true, for if $\text{how } x \text{ looks} = \text{how } y \text{ looks}$, by Leibniz's Law, for any F , x looks F if and only if y looks F ; there is no sorites paradox because the sorites argument merely shows that there can be unnoticed differences in looks. On the appearance of sameness reading, observation indeed shows that successive members of the sorites series look exactly alike, but the sorites argument then shows that Phenomenal Sorites is false. Although the premises ' x and y are exactly alike' and ' x is red' jointly entail the conclusion ' y is red', it does not follow that the premises ' $\text{it looks as though } (x \text{ and } y \text{ are exactly alike})$ ' and ' $\text{it looks as though } (x \text{ is red})$ ' jointly entail the conclusion ' $\text{it looks as though } (y \text{ is red})$ '; the sorites argument shows that there are counterexamples—if we carefully avoid illicitly switching contexts in verifying the premises. In the jargon, the sentential operator ' $\text{it looks as though}$ ' does not satisfy multi-premise closure; it is not a normal modal operator. Thus the reading on which Phenomenal Sorites continues to seem trivially true is not the reading on which we know that it is false. This example too fails to help Boghossian.

Of course, Boghossian could switch to a different paradox-threatening principle not subject to the same problem, for example:

(Sorites2) If two things are visually indiscriminable, then if one looks red, so does the other.

Boghossian might claim that Sorites₂ seems true to paradox-savvy subjects although they have no disposition to believe Sorites₂. That description does not ring true. I know that Sorites₂ is false, because it generates a Sorites paradox (on a suitable reading). The only sense in which it seems true to me is that I can still feel a slight temptation to assent to Sorites₂, which I am in no danger of giving way to. In other words, I still have a slight disposition to believe Sorites₂, which I easily override. Elsewhere, I have explained how such a residual temptation could result from a cognitively efficient, imperfectly reliable heuristic built into our psychology (Williamson 2020, pp. 63-67). Boghossian says nothing to convince me that the psychological state he has experienced is anything more than that.

In short, Boghossian's argument for consciously accessible pre-doxastic intellectual seemings fails. The next section considers some consequences of the justificatory role he assigns to such seemings.

4. *Consistent Nazis*

Boghossian sums up his theory of intuitions (intellectual seemings) by accepting three principles and rejecting two others. He accepts A, B, and C:

- A. Intuitions justify independently of background beliefs.
- B. Intuitions provide a priori justification.
- C. Intuitions sometimes provide such strong justification that they can defeat deeply entrenched beliefs.

He rejects D and E:

- D. Intuitions are infallible or indefeasible.
- E. If p is supported by intuition, it cannot also be supported by argument.

His account of perceptual seemings is similar, except of course that the justification they supposedly provide is a posteriori, not a priori. He explains the strength of the justification in C by adding that intuitions often present propositions as *obviously* necessarily true; similarly, he holds, perceptual seemings can present propositions as obviously true.

Such an internalist approach to justification faces the natural challenge that it is apt to count as justified the nastiest bigoted beliefs, when based on bigoted intuitions with the same content. Since Boghossian treats justification by intuition as defeasible, he allows such a justification to be defeated by a contrary intuition, so the most pressing form of the problem for him is the consistent bigot. In our book, I raised the old problem for internalist views of the consistent Nazi (Boghossian and Williamson, 2020, pp. 213, 238; see also Littlejohn, 2014, Srinivasan, 2020, and Williamson, 2019). The paper under discussion contains a more developed response than he provides in the book.

Boghossian's opening move is to question the feasibility of a consistent Nazi:

Any actual Nazi would have many defeaters for the odious belief that non-Aryans should be killed whenever possible. For example, any such person is likely to believe that human life is precious; that many people he admires and loves, some of whom may be friends or family, are non-Aryans; that you can't blame people for having features that are not under their control; that many non-Aryans have, throughout history, produced many things of great value; and so on and so forth. All of these beliefs would be in considerable tension with his nasty view.

That is over-optimistic. A Nazi may easily believe that *not* all human life is precious; only Aryan life is. He may convince himself that all his friends and family, all the people he admires or loves, and all those who, throughout history, have produced things which are *really* of great value, are *really* Aryans. He may agree that you can't blame people for having features that are not under their control, but regard the extermination of non-Aryans as not a matter of blame but more akin to the extermination of a virus. Such a Nazi is far from unimaginable. Boghossian's attempt to dismiss the spectre of the consistent Nazi is quite unconvincing.

Boghossian toys with the idea that he need only defend the weaker claim that the (inconsistent) Nazi's 'bigoted intuition provides him with *some* prima facie justification for his bigoted view'. That sets the bar too low. First, he has not succeeded in taking the consistent Nazi out of the argument. Second, even an inconsistent Nazi may have bigotry-inspired intuitions which present nasty propositions as *obviously* true, the key feature in his explanation of how intuitions can 'provide such strong justification that they can defeat deeply entrenched beliefs' (principle C above); in such cases, the inconsistent Nazi's bigoted belief may have all things considered justification, by Boghossian's standards.⁴

In the end, Boghossian adopts a non-conciliatory strategy, applicable even to consistent Nazis, whom he allows for the sake of argument and compares to brains in vats. On an internalist view, the usual sort of brain in a vat is *fully justified* in believing that it has hands, so Boghossian's account implies that a consistent Nazi is fully justified in believing that non-Aryans should be killed whenever possible. Naturally enough, he accepts such consequences of his internalist approach: 'Given that a priori belief can be fallible, allowing false intuitions to provide prima facie justification for false moral beliefs seems like the correct outcome'.

There is a suspicion of prejudice-laundering about such internalism. An unconscious prejudice that P manufactures an intuition which presents the proposition that P as true, even obvious, which in turn justifies the belief that P. The internalist norm of justification makes its entrance too late to do much more than tidy up the status quo. An exclusive focus on what is consciously accessible misses most of the cognitive action.

Unfortunately, 'justification' has become such a technical term in epistemology that its wider connections can be hard to awaken. Perhaps the best way of doing so is by connecting the justification of *belief* to the justification of *action* (Williamson 2019). After all, beliefs are there to be acted on, as noted earlier. It would be strange if norms of belief were *not* connected to norms of action.

Boghossian hastily moves to block the connection. He suggests that 'being justified in what to *believe* may be (as the internalist maintains) a function of the agent's *perspective*, in a way in which what the agent has reason to *do* need not be'. To motivate the asymmetry, he

adds ‘One type of reason, after all, has to do with how to change one’s mind, whereas the other has to do with how to change the world’, and comments ‘It wouldn’t be surprising if the former type of reason were more perspective-dependent than the latter’. The motivation fails. For a start, one’s mind is *part* of the world, so changing one’s mind *is* changing the world. Moreover, many decisions about changing one’s mind are clearly decisions about what to do: for example, whether to go to university.

Boghossian considers a candidate principle linking the justification of belief to the justification of action:

(JJ) If an agent justifiably believes that she should A, she would be justified in A’ing.

He objects to JJ, with reference to Feldman (1988):

For example, suppose that what action you are justified in performing is a matter of which action would have the best consequences. You may be perfectly justified in believing that A would have the best consequences, and so justified in believing that you should A, even though A is not the best action under the circumstances and so that [sic] you are not justified in A’ing.

This objection is under-motivated. It depends on choosing a perspective-dependent standard for the justification of belief but a perspective-independent standard for the justification of action, with no apparent reason for the asymmetry. One could just as easily have made trouble for JJ by choosing a perspective-*independent* standard for the justification of belief (say, truth) but a perspective-*dependent* standard for the justification of action (say, maximizing *expected* utility). If such asymmetries are allowed, no wonder (JJ) can be made to fail. The pick ‘n mix objection is too cheap.

Given that there are perspective-independent and perspective-dependent standards for both belief and action, friends of JJ will presumably defend it either for perspective-independent standards for both belief and action or for perspective-dependent standards for both. On the perspective-independent path, it is much more plausible that if the proposition that she should A is objectively the best thing for the agent to believe, then A’ing is objectively the best thing for her to do. Analogously, it is also much more plausible that if the proposition that she should A is the best thing for her to believe given her evidence, then A’ing is the best thing for her to do given her evidence.

Of course, the English language could have evolved so that the word ‘justification’ was most naturally read in the perspective-dependent way when applied to belief but in the perspective-independent way when applied to action, or vice versa. But that asymmetry would be of little philosophical interest without some deeper reason for it, which is not in sight.

When Boghossian concludes his brief discussion of links with the justification of action, he is cautious:

I don’t deny that there are many subtle issues that remain [...]; my only point is that we cannot automatically assume that the epistemic internalist is committed to something like JJ, and so cannot automatically assume that the putative full justification of [the Nazi’s] odious belief will translate into a justification for his *acting* on his odious belief.

He is surely right that the connections are not automatic. Nevertheless, they are plausible.

First, read both ‘justifiably’ and ‘justified’ as perspective-independent. Then if the agent justifiably believes that she should A, presumably she *should* A, and so presumably would be justified in A’ing. That supports JJ on a perspective-independent reading.

Alternatively, read both ‘justifiably’ and ‘justified’ as perspective-dependent. Then if the agent justifiably believes that she should A, presumably her perspective indicates that she *should* A, so presumably she would be justified in A’ing. That supports JJ on a perspective-dependent reading.

Neither argument is conclusive; both are natural. The second is more relevant to epistemic internalists: they are interested in a perspective-dependent reading of ‘justifiably’ for belief, so they should also be interested in a perspective-dependent reading of it for action. Indeed, the general internalist idea that it is unfair to judge agents by what they lack conscious access to is equally applicable to belief and action.⁵ If internalism vindicates nothing like JJ, and is forced to treat the justification of belief and action asymmetrically, that suggests something badly wrong with the motivation for internalism.

To put the point another way, consider an agent whose beliefs all strongly favour A’ing right now, but who refrains from A’ing. Presumably, that combination together is in no sense justified; something must be wrong with it. Hence, in any sense in which all her beliefs are justified, that justification should extend to *not* refraining from A’ing. For, presumably, justification is something like consistency with the relevant norms and conditions, whatever they are, whether internalist or externalist. Hence, for any X and Y, if X is justified, then X is consistent with the relevant norms and conditions, so either X & Y is consistent with the relevant norms and conditions or X & not-Y is consistent with the relevant norms and conditions, so either X & Y is justified or X & not-Y is justified, so if X & Y is not justified, then X & not-Y is justified. Thus something like JJ is hard to avoid, for both internalists and externalists.

Let Boghossian* be a philosopher who accepts both Boghossian’s general internalist approach and his specific account of the justification of belief, except that Boghossian* (a) accepts that consistent Nazis are possible and (b) accepts JJ on Boghossian’s understanding of the justification of belief and an analogously internalist understanding of the justification of action. Following Boghossian, Boghossian* calls the consistent Nazi ‘Hans’ and accepts (1):

- (1) Hans justifiably believes that he should kill non-Aryans whenever possible.⁶

An instance of JJ is (2):

- (2) If Hans justifiably believes that he should kill non-Aryans whenever possible, he would be justified in killing non-Aryans whenever possible.

Boghossian* applies modus ponens to (1) and (2) and accepts (3) on that basis:

- (3) Hans would be justified in killing non-Aryans whenever possible.

It is hard to regard (3) as a morally plausible conclusion. Indeed, even *trying* to regard (3) as morally plausible feels like an attempt to switch off one's own moral sensibility.

Could Boghossian* play down the moral significance of (3) by saying that it uses 'justified' in a merely technical sense, devoid of its ordinary normative connections? That would be quite at odds with the usual attitude of Boghossian and other internalists to matters of justification. They treat the justification of belief as the central normative issue in epistemology. That is the starting-point for their treatment of epistemic justification. Since Boghossian* models his understanding of the justification of action on Boghossian's understanding of the justification of belief, he treats the justification of action as the central normative issue in the philosophy of action. Thus Boghossian* is in no position to dismiss the moral significance of (3).

Presumably, Boghossian does not want to be Boghossian*. How can he avoid it? He must avoid either (a) or (b) above. His paper offers some resistance to both (a) and (b). To avoid (a), he must reject the claim that consistent Nazis are possible. He does indeed resist it, but, as we have seen, his reasons for doing so are underdeveloped and unpersuasive. To avoid (b) convincingly, Boghossian must show that either no reasonable understanding of the justification of action can be modelled on his understanding of the justification of belief, or no such analogous understandings of the justification of belief and the justification of action validate JJ. But if no reasonable understanding of the justification of action can be modelled on his understanding of the justification of belief, that looks like a failure of his account of the justification of belief to generalize properly, and so a threat to his internalist programme. Alternatively, if analogous understandings of the justification of belief and the justification of action can be developed along his lines, he has given no grounds to doubt that they will validate JJ, or something like it.

My remarks have been somewhat sketchy and programmatic, but that is a reflection of the sketchy and programmatic nature of Boghossian's response to the threat of a consistent Nazi. Adverting to the threat, he writes sarcastically of 'the question whether a philosophy that is friendly to intuitions inexorably leads to National Socialism', and of his hope to have shown that you can endorse the commitment to 'allowing false intuitions to provide prima facie justification for false moral beliefs' 'without fearing that it will immediately lead to the depredations of the Third Reich'. As far as I know, no one has those hyperbolic fears. No one is suggesting that Boghossian's theory of justification entails Nazi moral theory (unless his theory is inconsistent). Nor is anyone suggesting that his publications have much political influence. Rather, the point is that 'Hans would be justified in killing non-Aryans whenever possible' is obviously false, and any theory in danger of commitment to it has got hold of the wrong end of the stick.^{7,8}

Note

1 Boghossian leaves the nature of presenting obscure. Is ‘true’ meant to have some special magic, or does presenting the proposition that *a* is F just boil down to presenting *a* as F? If an image of a certain politician in prison irresistibly comes to my mind, is the proposition that he is in prison thereby presented to me as true?

2 The term ‘feeling of knowing’ is widely used in psychology, though mainly applied to examples such as feeling ‘It’s on the tip of my tongue’. However, it is widely accepted that there is a much more widespread phenomenon of metacognitive feelings, which play a functional role in cognition. Thus a feeling of fluency in processing is some indication of reliability, and so grounds confidence, while a feeling of disfluency prompts further investigation (Alter and Oppenheimer (2009), Alter, Oppenheimer, and Epley (2013); for epistemological discussion see Reber and Unkelbach (2010) and Nagel (2016)). Cognitively available markers of level of confidence associated with contents facilitate the integration of information from different sources, both within the general-purpose workspace of consciousness (Shea and Frith (2019)) and socially, between individuals in linguistic communication (Bahrami et al. (2010), Fusaroli et al. (2012)). Terms such as ‘see’ and ‘know’ are themselves used as such linguistic markers (Fusaroli et al. (op. cit.)). Shea and Frith (op. cit.) propose that every content in consciousness must carry such a metacognitive marker of level of confidence. Obviously, a feeling of knowing of that kind is not factive: one can have it about a falsehood. But these metacognitive feelings cannot play the role Boghossian wants seemings to play, for they are by no means pre-doxastic: they concern a prior judgment or estimate, made pre-consciously. In any case, a version of internalism which assigned the burden of epistemic justification to such metacognitive feelings would still be vulnerable to the arguments of section 4.

3 The authors of the article on which Srinivasan is relying warn ‘Whether flies are actually “susceptible” to the illusion is, of course, a logically separate and still open question’ (Geiger and Poggio, 1975, p. 480). There is some evidence that bees, which are cognitively more complex than flies, take in visual patterns sequentially rather than synchronically (Nityananda, Skorupski, and Chittka, 2014; Chittka, 2017).

4 See Rosen (2008) for discussion of what is required for defeat.

5 For related discussion see Huemer (2006), Gibbons (2013), Kiesewetter (2016), Way and Whiting (2016), and Lord (2018).

6 For present purposes, ‘should’ can be read as an extremely strong deontic modal, because for the sake of the example we can stipulate that Hans has correspondingly extreme beliefs. The more strongly we read ‘should’, the harder JJ is to dispute, because we have merely strengthened its antecedent.

7 Boghossian (2021) also contains a few critical remarks about passages he quotes from my contributions to Boghossian and Williamson (2020). Proper consideration of them would distract from the main line of argument in the present paper; in each case, a careful reading of the quoted passages will show that they already provide the basis for answers to his criticisms.

8 I am very grateful to Clayton Littlejohn, Daniel Kodsi, and Jennifer Nagel for detailed comments on a draft of this paper.

References

- Alter, A., and Oppenheimer, D. (2009). Uniting the tribes of fluency to form a metacognitive Nation. *Personality and Social Psychology Review*, 13, 219-235.
- Alter, A., Oppenheimer, D., and Epley, N. (2013). Disfluency prompts analytic thinking —But not always greater accuracy: response to Thompson et al. (2013). *Cognition*, 128, 252-255.
- Bahrami, B., Olsen, K., Latham, P., Roepstorff, A., Rees, G., and Frith, C. (2010). Optimally interacting minds. *Science*, 329, 1081-1085.
- Boghossian, P., and Williamson, T. (2020). *Debating the a priori*. Oxford: Oxford University Press.
- Boghossian, P. (2021). Intuition and a priori justification. In D. Dodd and E. Zardini (Eds.), *Beyond sense? New essays on the significance, grounds, and extent of the a priori*. Oxford, Oxford University Press, forthcoming.
- Chittka, L. (2017). Bee cognition. *Current Biology*, 27, R1049-R1053.
- Feldman, R. (1988). Epistemic obligation. *Philosophical Perspectives*, 2, 235-256.
- Feng, L., Chouinard, P., Howell, T., and Bennett, P. (2017). Why do animals differ in their susceptibility to geometrical illusions? *Psychonomic Bulletin and Review*, 24, 262-276.
- Fusaroli, R., Bahrami, B., Olse, K., Roepstorff, A., Rees, G., Frith, C., and Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science*, 23, 931-939.
- Geiger, G., and Poggio, T. (1975). The Müller-Lyer figure and the fly. *Science*, 190, 479-480.
- Gibbons, J. (2013). *The norm of belief*. Oxford: Oxford University Press.
- Huemer, M. (2006). Phenomenal conservatism and the internalist intuition. *American Philosophical Quarterly*, 43, 147-158.
- Kiesewetter, B. (2016). You ought to ϕ only if you may believe that you ought to ϕ . *Philosophical Quarterly*, 66, 760-782.
- Littlejohn, C. (2014). The unity of reason. In C. Littlejohn and J. Turri (Eds.), *Epistemic norms: new essays on action, belief, and assertion* (pp. 135-154). Oxford: Oxford University Press.
- Lord, E. (2018). *The Importance of being rational*. Oxford: Oxford University Press.
- Millar, S., and Al-Attar, Z. (2002). The Müller-Lyer illusion in touch and vision: implications for multisensory processes. *Perception and Psychophysics*, 64, 353-365.
- Mountjoy, P. (1958). Effects of exposure time and intertrial interval upon decrement to the Müller-Lyer illusion. *Journal of Experimental Psychology*, 56, 97-102.
- Nagel, J. (2016). Knowledge and reliability. In B. McLaughlin and H. Kornblith (Eds.), *Goldman and his critics* (pp. 237-256). Hoboken, NJ: Wiley.
- Nakamura, N., Fujita, K., Ushitani, T., and Miyatani, H. (2006). Perception of the standard and the reversed Müller-Lyer figures in pigeons (*Columba livia*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, 120, 252-261.
- Nityananda, V., Skorupski, P., and Chittka, L. (2014). Can bees see at a glance? *Journal of Experimental Biology*, 217, 1933-1939.
- Pepperberg, I., Vicinay, J., and Cavanagh, P. (2008). Processing of the Müller-Lyer illusion

- by a Grey parrot (*Psittacus Erithacus*). *Perception*, 37, 765-781.
- Predebon, J. (1998). Decrement of the Brentano Müller-Lyer illusion as a function of inspection time. *Perception*, 27, 183-192.
- Predebon, J. (2006). Decrement of the Müller-Lyer and Poggendorff illusions: the effects of inspection and practice. *Psychological Research*, 70, 384-394.
- Reber, R., and Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, 1, 563-581.
- Rosen, G. (2008). Kleinbart the oblivious and other tales of ignorance and responsibility. *Journal of Philosophy*, 105, 591-610.
- Rudel, R., and Teuber, H-L. (1963). Decrement of visual and haptic Müller-Lyer illusion on repeated trials: a study of crossmodal transfer. *Quarterly Journal of Experimental Psychology*, 15, 125-131.
- Shea, N., and Frith, C. (2019). The global workspace needs metacognition, *Trends in Cognitive Science*, 23, 560-571.
- Srinivasan, A. (2020). Radical externalism. *Philosophical Review*, 129, 395-431.
- Srinivasan, M. (1993). Even insects experience visual illusions. *Current Science*, 64: 649-655.
- Way, J., and Whiting, D. (2016). If you justifiably believe that you ought to ϕ , you ought to ϕ . *Philosophical Studies*. 173: 1873-1895.
- Williamson, T. (1990). *Identity and discrimination*. Oxford: Blackwell.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Williamson, T. (2019). Morally loaded cases in philosophy. *Proceedings and Addresses of the American Philosophical Association*, 93, 159-172. Reprinted in Williamson 2021.
- Williamson, T. (2020). *Suppose and tell: the semantics and heuristics of conditionals*. Oxford: Oxford University Press.
- Williamson, T. (2021). *The philosophy of philosophy*, 2nd ed. Oxford: Wiley-Blackwell.
- Williamson, T. (2022). Justifications, excuses, and sceptical scenarios. In F. Dorsch and J. Dutant (Eds.), *The new evil demon: new essays on knowledge, justification, and rationality*, forthcoming. Oxford: Oxford University Press.