

Siegel and the impact for epistemological internalism

Matthew McGrath

Published online: 6 December 2012
© Springer Science+Business Media Dordrecht 2012

If the preformationist “sees” tiny embryos when examining cells under the microscope because he wants to confirm his theory, it can seem that his “seeing” them to be there—his experience as of embryos being there—cannot justify him in believing they are there. If Jack looks angry to Jill because she unjustifiably expects him to look that way, how can she gain justification to believe he is angry from the experience? One certainly wants to say that there is something defective—epistemically defective—with the perceptual beliefs formed in these and the other cases Siegel gives. It can seem plausible to think, with Siegel, that these beliefs are unjustified, because of the bad etiology of the experiences on which they are based. Such etiologies might seem to downgrade the justificatory force of the experience. But what could explain why there is a justificatory downgrade in these sorts of cases but not in other cases in which experiences have an epistemically unfortunate etiology, e.g., in hallucination and in skeptical situations such as the brain in a vat or the demon victim? If the only options are an externalism which treats all these beliefs alike as unjustified and an internalism which treats them all as justified, perhaps we should learn to live with the internalism.¹

If Siegel is right, these do not exhaust the options. The etiologies of experiences in the cases of bad cognitive penetration seem importantly different from those in hallucination and other skeptical situations. The etiologies of the former are irrational in a way that those of the latter are not. Roughly speaking, an experience’s

¹ Externalists may cry foul. I have not mentioned actual-world reliabilism or proper-functionalism. I cannot consider here in any detail the merits of these and other externalist accounts of how subjects’ beliefs could be justified in skeptical cases. I will only say that it seems quite implausible that the reason why the demon victim’s perceptual beliefs are justified has to do with facts about our world or facts about our design or evolutionary past (think Swampman). For an externalist defense of the claim that etiology can affect an experience’s justificatory force see Goldman (2009, Sect. VI).

M. McGrath (✉)
University of Missouri, Columbia, MO, USA
e-mail: mcgrathma@missouri.edu

having an irrational etiology is a matter of that etiology essentially involving processes that resemble bad inference, processes which as she says may unfold beneath first-person radar, in the basement of the mind. If it is the irrationality of the experience's etiology which explains the downgrade in justificatory force in the cognitive penetration cases, then we can see this as an "internal" defect, as opposed to "external" defects such as unreliability, insensitivity, and the like found in the skeptical cases. This would be an important result. It would show that reasons to back away from the strong claim that all experiences as of *p* prima facie justify one in believing that *p* need not translate into reasons to give up internalism. More precisely, such reasons need not translate into reasons to give up mentalistic internalism or mentalism, i.e., the thesis that which propositions one is justified believing is determined by the totality of one's mental states.²

Compare two brains in vats.³ One has an experience as of something being a gun because of the normal use of one's perceptual identificatory skills (skills which, if one were embodied would produce normal perceptual knowledge) and the other has an experience as of something being a gun due to fear. Both form beliefs that there is a gun before them based on these experiences. Both BIVs' resultant beliefs are externalistically compromised: unreliably formed, unsafe, insensitive, etc. Yet the belief of the second BIV, whose experience is fear-induced, seems epistemically compromised in a way that the first's is not. It is not at all unnatural to describe the epistemic defect by using the language of justification: the fearful BIV's belief is unjustified, irrational, unreasonable. Siegel gives us a way to see how this description could be correct: the etiology of the fearful BIV's experience is irrational, thus making her belief based on it unjustified; but not so for the "normal" BIV. There is something like a process of a bad reasoning underlining the experience of the fearful BIV but not that of the normal BIV's. The explanation is fully consistent with mentalism: it is because of certain mental factors in the etiology of the experience that the fearful BIV's experience (and not present in the etiology of the normal BIV's experience) that the fearful BIV's experience cannot justify belief based on it.

However, even if Siegel succeeds in showing how an internalist could agree that in some cases an experience's etiology can downgrade its justificatory force, internalists may see no reason to agree. Yes, they *could* consistently accept the Downgrade Principle,⁴ but they might remain unmoved by the cases presented to support it. They ought to agree that there is something epistemically defective about the subjects' perceptual beliefs in the cases Siegel considers, and they ought to agree that the defects aren't simply reducible to unreliability or other externalistic faults (since they can be found just as well in some but not all BIVs). But they may hope to chalk up the defects to epistemic irresponsibility or epistemic vice, rather than to justification.

² For a defense of mentalism, see Conee and Feldman (2001). In what follows, I use 'internalism' to refer to mentalism rather than to any thesis about introspective accessibility of justifiers.

³ Here I borrow from an example used by colleague Peter Markie.

⁴ The Downgrade Principle holds that "the justification an experience provides for its contents falls below the baseline," i.e., cannot justify beliefs in those contents (p. 7). Siegel considers two ways to develop this principle, one which takes the principle as a principle about propositional justification (the Propositional Downgrade Thesis) and the other which takes it as a principle about doxastic justification (the Doxastic Downgrade Thesis). I will treat these as interchangeable.

My aim is to defend Siegel's Downgrade Principle, but in a somewhat different way than Siegel herself does. I rely on the very framework—evidentialism—that provides internalist critics their best argument against the principle. I begin by considering how an internalist might argue against Siegel's principle, focusing on a very plausible evidentialist argument against it. In §2, I raise doubts about the prospects for Siegel's own theoretical argument for the principle. In §3, I argue that the very evidentialism that the internalist might hope to use *against* the principle in fact supports it: when experiences are badly based on other experiences, the output experiences do not count as evidence at all. Finally, in §4 I discuss the question of which sorts of etiologies for experiences preclude them from being evidence, suggesting that the etiologies which seem most clearly to lead to justificatory downgrades are those involving something like inference, which I call "quasi-inference." My "quasi-inference" proposal amounts to a way of spelling out what it takes for an experience's etiology to count as irrational, in such a way that we can see why experiences with such etiologies are not evidence for the subject.

1 Two internalist arguments against the Downgrade Principle

The internalist can do more than merely suggest that that the subjects in Siegel's cognitive penetration cases might be justified; he can argue against Siegel's Downgrade Principle. I'll consider two arguments, the second of which I think is more promising. The first, though, is a very natural one.

The first goes like this. Suppose something looks red to you, and suppose you have no reason to think it isn't and no reason to think it would look red to you even if it wasn't. What are you supposed to think? Are you to believe it *isn't* red? That would be very unreasonable. Are you to withhold judgment on whether it is red? But there it is, plainly looking red! No, the only reasonable attitude to have is belief: belief that it is red. Generally, when you have an experience as of *p*, and you lack defeaters, what are you supposed to do doxastically? Disbelief isn't reasonable, nor is withholding judgment, only belief that *p* is.⁵ Notice the generalization applies to *all* cases of experiencing as of *p* without defeaters, and so it applies to cases of checkered experiences. Thus, the Downgrade Principle is false.

I feel the pull of this sort of "what am I supposed to think?" argument. But this sort of argument is *very* powerful, too powerful. Suppose you are considering the fact that your child plays the piano very well, better than anyone you can now think of in her teacher's piano studio, and suppose you believe you've remembered all the children in the teacher's studio. In fact you are just not thinking about Maggie, who is also a child in the teacher's studio and who plays better than your child. If you were to think harder, you'd of course remember Maggie and remember just how well she plays. As it stands, you don't. Now, what are you supposed to think in the situation? You do have the knowledge that Maggie plays better than your child but you aren't bringing that knowledge to mind. Given that you're not bringing it to mind and that the evidence you

⁵ If we understand justification as a sort of epistemic permission to believe, then even if we supposed withholding was permissible here, surely belief is as well—so it might be argued.

are bringing to mind supports the proposition that your child is the best in the studio, it can seem that you are supposed to believe that your child plays better than anyone else in the studio. But even still, I take it that you are not justified in believing your child is the best in the studio, because you have strong counterevidence.

Or consider an even more extreme example, which I describe schematically. Suppose you recall lots of evidence against P. Still, you manage to believe P. Now, let's ask: given that you do believe P, what are you supposed to think about the disjunction P-or-Q? True, you have evidence against P, and (let us also suppose) evidence against Q. But you're not weighing that evidence correctly. Given how you're weighing it the evidence and that you do believe P, wouldn't it be incoherent to disbelieve or withhold on P-or-Q? Surely, the only thing it could make sense for you—in your situation—to believe P-or-Q. However, here even more clearly than in the piano case, I take it that you aren't justified in believing P-or-Q, because you have clear evidence against P and against Q.

Intuitions about what one is supposed to think *given* other mental states one is in are therefore not a reliable guide to facts about what one is justified in believing.⁶ This first internalist argument against the Downgrade Principle fails.

The second internalist argument is more compelling. It still relies on intuitions about what it is reasonable to think—what one is supposed to think—but not just given various mental states one happens to be in, but given one's *evidence*. It relies, in other words, on evidentialist principles.⁷ Here is the argument. In cases of normal perception as well as skeptical cases, experiences as of P justify belief that P. The reason experiences justify in these cases is that, generally, when you have an experience as of P, you thereby have good evidence or reason to believe P; and so if you lack defeaters the only attitude it is reasonable for you to have is *belief* that P. If this is generally true, then it is true in Siegel's cases. So, in Siegel's cases, the attitude it is reasonable to have toward P in light of the fact that one has the experience as of P and lacks defeaters is belief. Siegel's cases must be cases of justified belief if the normal and skeptical cases are cases of justified belief, which they are.

We will consider responses to this argument in the next two sections. Let me first try to improve even further the internalist's overall case against Siegel by considering how the internalist might explain what is going wrong in Siegel's cases.

The internalist might assimilate Siegel's cases to cases like the following:

Biased Search. John, unbeknownst to himself, searches only for confirming evidence for his being a good teacher, and he finds it; now, given that he has it, and knows nothing of the biased process he used, what's he supposed to do? Answer: believe he is a good teacher.

Fabricated Evidence. Mary, unbeknownst to her at the present time, created evidence that she was a good teacher, by creating TEQs to confirm it and writing glowing reviews on Rate My Professors. At the present time, she has

⁶ They may indicate something important about principles of justification, though, if such coherence requirements are, so to speak, shadows of principles of justification. I cannot discuss the issue further here.

⁷ For an explanation and defense of evidentialism, see Conee and Feldman (2004).

all this strong evidence and doesn't know its origin. What is she supposed to do? Answer: believe she is a good teacher.

The argument might be: in Biased Search and Fabricated Evidence, the subject employs a faulty evidence-gathering procedure—the procedure is at the very least epistemically vicious, and perhaps also irresponsible—but at the end of the process the subject does have good evidence for the conclusion and no good evidence against it; and surely one ought to form one's doxastic attitudes so as to fit the evidence one has at the time; thus, John and Mary are justified in their beliefs. The same goes for the subjects in Siegel's examples. They do have evidence—the evidence of their senses—for their beliefs. However defective the process through which they got the evidence, it is still *evidence*, and given that it is unreasonable to believe contrary to one's evidence, they are justified in believing what they do.⁸

2 Siegel's argument for the Downgrade Principle

Siegel gives an argument for justificatory downgrades which does not rest on intuitions about her cases. The argument is based on a parallel with belief. An unjustified belief with an irrational etiology can make further beliefs based on it unjustified. Through an unjustified belief the irrational elements in the belief's etiology can tarnish further beliefs. In the absence of any reason to think that beliefs can do this only by virtue of some feature which they do not share with experiences, we ought to conclude that experiences with similar etiologies can make beliefs based on them unjustified. That is, we ought to accept the Downgrade Principle. Siegel considers, one by one, various features beliefs can have which experiences cannot: beliefs can be justified or unjustified (irrational or irrational), experiences cannot; beliefs can be dispositional, experiences can only be occurrent, beliefs can be formed as a result of explicit reasoning, experiences cannot, etc. She argues that, for each of these properties, we can find cases in which beliefs with irrational etiologies lack the property in question but still make beliefs based on them unjustified. Thus, none of these properties can account for why beliefs manage to serve as conduits through which the irrational elements in their etiologies make further beliefs based on them unjustified.

I am unclear, though, about how Siegel can rule out the hypothesis that when an unjustified belief with an irrational etiology makes a belief based on it unjustified it does so only because it itself is unjustified. The problem is that when a belief is based on an irrational etiology, it has to be unjustified. So, how can we hold fixed the irrational etiology while manipulating the belief's status as unjustified to see if it still makes beliefs based on it unjustified? We can't. So, we can't rule out the possibility that beliefs with irrational etiologies can tarnish beliefs based on them only because those tarnishing beliefs are themselves unjustified.

⁸ The internalist account given here of Biased Search and Fabricated Evidence is similar to that of Feldman (2000, Sect. D), with one exception: Feldman does not grant that the defects here count as epistemic. I would differ on this score. These defects alone seem to prevent the beliefs formed from being knowledge. I take this to be enough to qualify those defects as epistemic.

Siegel argues that there are cases in which beliefs are rational (justified) but manage to generate unjustified beliefs on their (partial) basis. If spiders frighten me, then if I see a spider and come to have a justified belief that there is a spider before me, this can generate, together with my fear, an unjustified belief that I'm in danger. But this isn't a case of a belief tarnishing beliefs based on it, because of its irrational etiology. It is only a case of a belief serving as an essential ingredient in an overall irrational etiology for another unjustified belief. We can grant that experiences sometimes do that: an experience as of a spider, together with a fear of spiders, can give rise to an unjustified belief that one is in danger. But that is no reason to think the experience as of a spider is itself epistemically downgraded: one would be still justified in believing the thing was a spider based on the experience.

The internalist, moreover, will mention another difference between experiences and beliefs. Experiences can be evidence or reasons for belief; beliefs cannot. When you have an unjustified belief that *p* with an irrational etiology, this by itself gives you no evidence for propositions supported by *p*. When you have an experience with an irrational etiology, this *does* give you evidence for *p*. Indeed, the internalist will say that it's the evidential power of experience that accounts for our intuitions that normal BIVs, hallucinators, etc. are justified in their perceptual beliefs.

3 An evidentialist defense of the Downgrade Principle

There are several ways one might try to attack the (evidentialist) internalist argument against the Downgrade Principle, which I discussed in §1. One could attack the underlying evidentialist assumption that having good undefeated evidence is sufficient for justification. One might argue that in Biased Search and Fabricated Evidence, one does indeed have undefeated evidence and yet one isn't justified in one's belief. Alternatively, one could accept the evidentialist assumption but argue that the range of defeaters must include not only what Siegel (2011) calls evidential defeaters but also propositional ones, including facts inaccessible to the subject. To give up the standard form of evidentialism is not necessarily to give up mentalistic internalism. So long as the propositional defeaters are facts about the subject's mind, either at the time or earlier, and so long as the failure to be justified in Biased Search and Fabricated Evidence can be explained in terms of the subject's mental history, we can retain mentalism (although it may prove difficult to do this in a non-ad hoc way).

However, it is worthwhile to ask whether the Downgrade Principle could be acceptable assuming the truth of standard evidentialism—with its narrow construal of defeaters and its insistence that the only evidence that matters is evidence had by the subject at the time.

Consider three of Siegel's cases: Anger, Preformationism, and Pliers.⁹ It is natural to see these cases as ones in which the subject's higher-level experiences are wrongly "based" on the lower-level experiences, due to the influence of a cognitive state. Jill's experience of Jack as angry seems to be based on an experience of his

⁹ I think the same analysis applies to her other examples, Ouija Board and Pessimism.

looking a certain way which in fact isn't a good indicator of anger for her. There is a gap here that is filled in, or leapt across, only by the expectation that Jack is angry, not by Jill's knowledge connecting the lower-level content with the higher-level one. The preformationist sees something of a certain shape, size and texture which in fact isn't a good indicator of an embryo, but which serves as his basis for seeing it as an embryo. Here again there is a gap between the shape, size and texture features and the embryo feature, and it is only the desire to see one's theory confirmed which "closes" the gap. The pliers have a certain color, shape, and size—a certain gestalt—which is not much of a good indicator of a gun, but nonetheless, on this basis, with the help of the prejudice rather than any background knowledge or perceptual ability, the subject sees the object as a gun.

In all these cases, it intuitively seems the subject is doing something very much like "jumping to conclusions" in having the higher-level experience based on the lower-level experiences. The jumping is not strictly speaking inference, since inference arguably connects only doxastic states. But it is inference-like. To confirm this, let us suppose that one transitioned from a perceptual belief with lower-level contents to perceptual belief with the higher-level contents, again because of prejudice. Moving from a belief that the object (i.e., the pair of pliers) has *that* gestalt to the belief that it is a gun is a case of jumping to conclusions. The gestalt is only weak evidence that the thing is a gun, but one's prejudice leads one to treat the weak evidence as good enough. Let us say that this sort of *basin*g of experiences on other experiences amounts to *quasi-inference*. It deserves this name because the same sort of basing, when it connects beliefs, counts as genuine inference. How could the fact that an experience is quasi-inferred bear on its status as evidence for the subject? I will give just a few brief suggestions.

First, what is quasi-inferred is not fundamental evidence; and so it cannot provide foundational justification. What provides foundational justification must be "handed" to us, rather than what we "make" of it. It must count as the world presenting itself to us in a certain way. This is not to say that cognitive processes do not mediate and help to explain our having the foundational evidence we do. But these processes cannot, I think, amount to the *subject's* inferring or quasi-inferring the foundational evidence. It is consistent with this to think that subpersonal processes, which might themselves count as inferential (or at least computational) help to explain how we have our foundational justification. Call experiences which are not quasi-inferred *receptive* experiences. The idea that there are receptive experiences is fully consistent with a Hemholtzian picture of perceptual experience as the result of inference—it is just that the inference is not inference or quasi-inference *by the person*.¹⁰

Second, even though nonreceptive experiences are not fundamental evidence, and so do not provide foundational justification, they could still serve as a conduit for such evidence, and so count as providing justification *if the quasi-inference is a good one for the subject*. The question of what makes a quasi-inference a good one for the subject might be usefully approached by asking what makes the corresponding inference between beliefs a good one. Suppose a birder sees a goldfinch in the winter. Inference from the belief that it has a certain color, shape,

¹⁰ The idea of quasi-inference in experience is discussed more fully in McGrath (forthcoming).

size (generally gestalt) to the belief that it is a goldfinch is a good one, given the birder's background knowledge of the appearances of goldfinches in winter and given knowledge that it is winter. If someone without this knowledge of the appearance of goldfinches in winter were to make the same inference, the inference would not be a good one for him. Similarly, whether a quasi-inference is a good one for the subject is a matter of whether the content of the input experiences, given the subject's background knowledge (whether it is knowledge-that or knowledge-how), sufficiently supports the content of the output experience.¹¹ On the account I am suggesting, if it does, the output experience justifies the subject in the belief.

This is just a sketch of a proposal, but I think it hints at a plausible way to see how there could be justificatory downgrades consistently with the truth of standard evidentialism. And it is very much motivated by Siegel's guiding idea that some etiologies count as rationally assessable and some do not, and it is the rationally assessable ones that can lead to downgrades when the inference-like processes involved are epistemically bad.

4 Beyond quasi-inference?

If one agrees that an experience's being the output of bad quasi-inference downgrades that experience's justificatory force because it prevents it from being evidence, we might wonder whether other "irrational" features of an experience's etiology could prevent it from being evidence and thereby downgrade it. That is, being badly quasi-inferred could be one sort of irrational etiology. If there are others, which we could see to prevent the experience from counting as evidence, but which would not apply to experiences in standard skeptical cases, then internalists would have reason to go beyond the quasi-inference account I've sketched in the previous section. Here I raise some questions about what these other features might be. I tentatively suggest that the plausibility of claims about epistemic downgrades is enhanced when there is something like bad reasoning underlying the experience, and thus we should be reluctant to go beyond the quasi-inference account.

One might think that if one's desire to believe a certain conclusion influences what the inputs are to processes of quasi-inference, biasing those inputs toward the desired outcome, this fact disqualifies the output experiences from being evidence. However, as Siegel notes in her (2011), this is too strong. Clearly, not just *any* sort of influence of inputs by such directional goals will disqualify the outputs from being evidence. If I want to believe the lights are on, as in Feldman's example (2000, pp. 671–672), this might lead me to turn them on, thus affecting my perceptual inputs and giving me genuine evidence that the lights are on. The influence of the directional goal must, at the very least, occur internally, i.e., after

¹¹ To preserve mentalism, we would need to insist that the background information needn't be knowledge in order to play its role in making the quasi-inference a good one for the subject. But I think this is plausible. Take a "normal" BIV who is justified in her belief that there is a gun before her. She is justified because her experience as of a gun is the output of a good quasi-inference. Although she doesn't count as knowing the appearance of guns, and perhaps doesn't count as knowing how to identify guns, she has mentalistic surrogates of these forms of knowledge. I assume here that knowledge is not a mental state.

the initial stimulation of sensory receptors. But even here there are difficulties. Suppose I want to believe that my child is smiling and this primes me to identify even the slightest smile where I otherwise would not. In this case, my experience as of the smile looks like genuine evidence. The same considerations, I think, apply to influence on experience by fears, prejudices, expectations, etc.

The influence seems worse when it can be seen as freely adding something to the processing which was not there to be extracted the first place. So, suppose (fancifully) that a directional goal could directly influence how much certain ganglion cells fire so as to produce experiences of yellow, where without this influence the experience would be of blue. Is the resulting experience evidence? One might claim that it is not. Here I'd tentatively suggest that the source of the worry is that one's experience is the result of something like practical reasoning. If a desire to believe something completely unrelated to yellowness somehow caused the spike in the firing of the ganglion cells, this alone wouldn't seem to disqualify the yellow experience from being evidence. (It might affect whether one can gain knowledge that something before one is yellow, of course.) But when the desire and the experience line up in a way that suggests reasoning, the intuition (for me) is stronger than this is not genuine evidence. If we could understand such a case as a case of practical reasoning, perhaps of the Aristotelian sort in which an "action"—in this case an experience—is the conclusion of the practical inference, we might expand our account of quasi-inference to cover it.

Another sort of etiology which appears in some sense irrational is one in which one's desire to believe a certain conclusion P affects one's attentional or pre-attentional selectional mechanisms in such a way that one disposes oneself to have P-friendly experiences and *not* to have P-unfriendly experiences. Siegel gives the example of white subjects' experiences as of the black faces being the only angry faces in a picture, due to their biased attention to black faces. One might think that the "only" experience is downgraded, even if the particular experiences as of the particular black faces being angry are not. If we could see the "only" experience as the outcome of something like bad inference from the particular experiences, we could see this as a case of bad quasi-inference. It appears there is a gap between the particular experiences and the "only" experience, which it is being bridged illegitimately.¹²

5 Conclusion

Siegel's key claim, that the irrationality of an experience's etiology can downgrade its justificatory force, seems to me correct and important. It shows us how we can accept mentalistic internalism while agreeing that the perceptual beliefs of subjects

¹² In personal communication, Siegel suggests that there might be cases of such selection effects in which the person is given substantial evidence of their ability to detect features in a set of racially diverse faces. If so, what exactly would be bad about the quasi-inference from the particular experiences to the "only" experience? Still, wouldn't the "only" experience be downgraded, and downgraded because it is not evidence? However, I would still be tempted to think that in such a case the "only" experience would fail to be evidence because it is not arrived at in the right way from the subject's fundamental experiential evidence.

whose experiences are the results of wishful thinking, prejudice, and fear are unjustified. Here I have offered an evidentialist defense of Siegel, one appealing to the idea that experiences do not count as evidence when they are based on bad quasi-inference. This defense also enables us to see why a promising internalist argument against the Downgrade Principle fails.

Acknowledgments Thanks to Richard Feldman and Susanna Siegel for helpful discussion.

References

- Conee, E., & Feldman, R. (2001). Internalism defended. In H. Kornblith (Ed.), *Epistemology: Internalism and externalism* (pp. 231–260). Oxford: Basil Blackwell.
- Conee, E., & Feldman, R. (2004). *Evidentialism*. Oxford: Oxford University Press.
- Feldman, R. (2000). The ethics of belief. *Philosophy and Phenomenological Research*, 60(3), 67–695.
- Goldman, A. (2009). Internalism, externalism, and the architecture of justification. *Journal of Philosophy*, 106(6), 309–338.
- McGrath, M. (forthcoming). Phenomenal conservatism and cognitive penetration: The ‘bad basis’ counterexamples. In C. Tucker (ed.) *Seemings and justification: New essays on dogmatism and phenomenal conservatism*. Oxford: Oxford University Press.
- Siegel, S. (2011). Cognitive penetrability and perceptual justification. *Nous*, 46(2), 201–222.
- Siegel, S. (2012). The epistemic impact of the etiology of experience. *Philosophical Studies*. doi: [10.1007/s11098-012-0059-5](https://doi.org/10.1007/s11098-012-0059-5).