

ERNEST SOSA

INTERNAL FOUNDATIONS OR EXTERNAL VIRTUES?¹

1. TWO PROBLEMS FOR INTERNALIST FOUNDATIONALISM

According to Larry BonJour, I am foundationally justified when my introspective belief seems to accurately characterize the conscious experience that it purports to describe, provided there is no special reason for suspecting error.² If such a *seeming* is to provide a proper foundation for justification, then presumably it is not itself something in need of justification. If it were itself a judgment that my belief is true, what would justify *this* judgment? This would launch a regress.

Accordingly, I will assume that the seeming is not constituted by a meta-judgment about the accuracy of one's introspective belief. It may involve rather just the content of the belief and its relation to the experience that it's about, except only for the proviso that there be no defeating reason in the offing. Briefly put, then, the view would be this:

We are foundationally justified if we undergo an experience with a certain content and believe we do, so long as there is no special reason to suspect error.

Only an epistemic reason would be relevant, presumably, not a pragmatic or moral or any other practical reason. So, the requirement, that one have no special reason to suspect error, amounts to a kind of "negative coherence" requirement: that is, one cannot at the time in question have within one's total mental state any elements that would strongly enough support the view that in believing <p> one is making a mistake.

BonJour and I agree that one's visual experience can have a character that one overlooks, because it is discernible only through careful attention, and perhaps only through careful counting. In such a case one will be "e-aware" ("experientially-aware") of that character, in the sense that one's experience does have that character as content. BonJour and I also agree that the subject may still overlook the relevant character of his experience even if he pays careful attention. We agree, moreover, that even careful attention is insufficient to yield justification. One might have a sensory experience with a certain content, believe that it has that content, and attend carefully to the relevant features of the experience, while failing to be justified in holding that belief.

What then must be required beyond one's careful attention to the relevant features of the experience? In his view, the subject must also *apprehend* the relevant congruence between the experience and the belief. However, this further apprehension is not a belief or a judgment or the intellectual application of a concept.

So, consider a horizontal array of 11 fat, black dots clearly visible against a pure white background, in an otherwise dark room, and please focus on your state of consciousness itself, on your visual experience as of seeing such an array. Suppose you are consciously convinced that the array has 11 members. Your state of consciousness at the time hence features the visual e-awareness of the 11-member array along with the conscious conviction that the array has 11 members. However, you may still fail to be justified in that belief. What accounts for this failure? BonJour proposes this: that you fail to "apprehend" the agreement or fit between the cardinality of the array and the cardinality attributed to it by the conscious conviction. But this failure is not just a failure to judge intellectually that the two fit. An apprehension is not just a belief or judgment; it is rather some sort of non-cognitive apprehension.

This proposal is rejected by Rich Feldman; he finds mysterious any such *non-cognitive* apprehension of a fit between an experience and a corresponding conviction, an apprehension

that is not a judgment. According to Feldman, the internalist should instead require not just that one undergo a sensory experience with a certain content, but that one “apprehend” that one’s experience has that content. He thus prefers to sandwich an apprehension between the experience and the belief, in order to account for when the belief is justified. This for him will be the important difference between an array with three dots and one with 11. In the former case there is a mediating, post-experience, pre-belief apprehension, whereas in the latter there is not. I confess to having no good grasp of any such mediating apprehension. I myself find this about as mysterious as any *sui generis* “apprehension” of congruence, between an experience and a conviction, one that is *neither* a belief or an inclination to believe *nor* just the sheer fact that the contents of one’s consciousness include both such an experience and also such a congruent belief.

Feldman appeals to recognitional concepts in an attempt to throw light on his sandwiched apprehensions. If I have understood properly, one must have a recognitional concept of a property if one is to apprehend that property in one’s experience. A property *can* figure in one’s experience even if one has no recognitional concept of it, but in that case one would *not* apprehend it.

That being so, why do we need the apprehension in addition to the recognitional concept? Why not say that a judgment about the character of one’s experience is justified if it is the exercise of a recognitional concept of such experience? Since the exercise of a recognitional concept is bound to be reliable, this will fit reliabilist intuitions nicely. Indeed, this would seem a version of reliabilism. Foundational empirical beliefs may then be said essentially to involve the exercise of recognitional concepts, and such exercise is bound to be largely reliable, by definition of recognitional concepts.

Suppose, on the other hand, that we back off and require only a supporting apprehension, without requiring that it always derive from the exercise of a recognitional concept. Such an apprehension would then mediate between the experience with a certain content and the forming of a belief that

one has such an experience. In simple cases of three speckles there is allegedly such a mediating apprehension, whereas in speckled hen cases of 11 or more speckles the apprehension is missing and with it the justification for the belief. And now we face this question: Are such apprehendings fallible? Invoking failsafe apprehendings would clash with the evident fallibility of our foundationally justified introspective beliefs. Indeed, as noted earlier, BonJour is himself committed to the fallibility of such foundational introspection, and Feldman does not demur.

Accordingly, suppose such apprehendings to be fallible. If so, how then are they epistemically helpful? Might they not be radically unreliable? If so, would they really help explain how, once we are up to 11 speckles or more, we forfeit epistemic justification? What if we *did* host apprehendings with cardinality that high, but these turned out to be radically unreliable? Would we then enjoy epistemic justification for our corresponding beliefs? Compare the case of Ramanujan. Is it not the amazing reliability of his intuitions that accounts for their probative force? What probative force could the intuitions of a Ramanujan have if they derived only from apprehendings that made his beliefs about as unreliable as ours? And the same would seem to hold for a Ramanujan of introspection as for a Ramanujan of mathematical intuition.

BonJour has now helpfully highlighted a question at the heart of the history of epistemology, one that is the main focus of his own work. The question is what reasons we have for our beliefs, both for our empirical beliefs and for our *a priori* beliefs. It is on this question that he takes externalism to be misguided. His is a “first-person” question, an answer to which will require the subject’s own access to the alleged reasons. No external reason will be one that the subject relevantly *has*. Reasons that one has must be relevantly *accessible*.

It seems to me that this question is indeed important for epistemology, and that BonJour has done well to focus on it. It would be nice to get a clearer view of what exactly is at issue, however, and why it is important. Accordingly, what follows will tackle the following questions: What are

foundational reasons, exactly, and why must these reasons be accessible?

Why not allow reasons why an agent should proceed a certain way, intellectually in forming a certain belief, or practically in opting for a certain choice: reasons that are there whether they are accessible to the agent or not? Can there not be reasons for a course of action that are beyond people's ability to discern or even understand? Surely there can. So why restrict the reasons that figure importantly in our normative disciplines to those accessible to the subject?

In evaluating an action we can be interested externally in what is good about that action, in what makes it a good thing to have performed. This will presumably involve an evaluation of the consequences of that action. And such consequences can of course affect how good or bad the action is, regardless of whether the agent knows about them.

In evaluating an action we can also be interested, however, in how well formed rationally it is, in its rational motivation. What reasons did the agent *have* for or against acting that way?

What then does it mean for the agent to *have* a reason? Is it just for there to *be* that reason? In one sense yes; an agent can have a reason for doing something even if he is unaware that he has it. But in another sense no; one really has only the reasons within one's awareness.

In evaluating an action motivationally what matters is not restricted to what reasons the agent *has*, in a sense compatible with his ignorance of those reasons and even with his inability to grasp them. Nor is our interest restricted to what reasons are accessible to the agent, even if he does not in fact access them. Nor is it restricted even to what reasons the agent has *and* are accessible to him, and are indeed *accessed* by him. In evaluating an action motivationally it matters also what reasons an agent acts *for*, what reasons actually motivate that action, and the quality of these reasons.

If we are interested in such evaluation of action, in how well-formed the relevant choice or intention may be, then we will be interested not so much in its consequences as in its

motives. And the same goes for epistemology. We may be interested in the rational evaluation of belief, in the *epistemic* rational evaluation of belief, in which case what matters is the reasons *for* which the believer believes as he does. Reasons for so believing, but inaccessible to the subject, will not bear on such evaluation (if we neglect negligence).³

With this in mind, let us return to the classical foundationalist account of our fundamental empirical reasons for belief. According to that account, such reasons reside in the empirical given, but how? Answer: we are supposed to believe with direct sensitivity to what is given in consciousness, with no dependence on inferential reasoning. Take a throbbing headache. Our reason for believing we have that headache is the fact that the headache is right there in our consciousness. The reason is just the headache. One is foundationally justified in believing that one suffers the headache simply because one does suffer it and believes one does, while there is no special reason to suspect error. Given our focus on the reasons that one has for one's beliefs, and the reasons for which one believes as one does, we must ask what constitutes one's reason for believing one has that headache. Is it the headache itself, is it one's undergoing that experience? Is that the positive reason one has, foundationally, for believing as one does, which rationally justifies one's belief, in the sense of rationally well-motivating the belief?

Problematically for this account, it seems to require infallibilism of foundational rational motivation. Take a case where one's belief is false but foundationally justified. How would we understand such justification? It could not be in terms of BonJour's traditional account, for this requires congruence between the experience and the foundationally justified belief, such that the incorrectness of that belief is necessarily ruled out. Recall: on that account one is foundationally justified in an introspective belief that one undergoes a certain experience when one *does* undergo the experience and believes one does while apprehending the experience-belief congruence, with no reason to suspect error. No introspective belief thus justified could possibly be false.

That being so, here is a new problem for internalist foundationalism: How is it to account for false foundationally justified belief? As far as I can see, this is an open question for classical internalist foundationalists.

It might be argued that classical foundationalism is concerned not with rationally well motivated beliefs, but with the issue of antecedent justification, with what reasons there are in favor or against believing propositions in certain large classes of traditional interest: about the external world, for example, or other minds, and the rest. But such an interest cannot plausibly stand alone. Our interest in such reasons is presumably ultimately motivated by our desire to form beliefs well in such domains. We are not just interested in the abstract question of whether there *are* such reasons and what they are, just out of curiosity. On the contrary, we are interested in how beliefs are formed well in those domains. So we are interested not just in what reasons there might be, but also in how we can be well guided by those reasons, so as to form beliefs well in the domains of interest. Accordingly, we would like to know what it is that renders beliefs well formed and well held. As we have seen, this creates a problem for classical internalist foundationalism, one whose solution remains to be seen.

We have considered two problems for classical internalist foundationalism: the problem of the speckled hen, and the problem of false foundationally justified belief. It seems to me that both problems argue in favor of a virtue epistemology, but arguing for this must be postponed to another occasion.⁴

2. REPLY TO JONATHAN VOGEL

Jonathan Vogel's astute comments raise two main issues. He first discusses my critique of epistemic internalism, arguing that Cartesian internalism survives that critique even if Chisholmian internalism does not. According to Cartesian internalism, epistemic justification derives from the subject's proper thinking, regardless of how well- or ill-circumstanced he may be. Chisholmian internalism, by contrast, requires epistemic justification to be reflectively accessible to the

believer in the armchair. This latter internalism is hence unable to allow time-involving memory and inference their proper roles as sources of epistemic justification. For, the presence and quality of such time-involving processes will not be accessible to the believer in the armchair at the later time when his belief is justified. However, they can still help constitute time-involving proper thought by that subject, so they can still help to determine or constitute his proper thinking across the relevant stretch of time. Unlike Chisholmian internalism, Cartesian internalism *can* therefore allow to such processes their proper role as sources of justification.

However, Vogel raises a problem for any fallibilist who attributes to memory any such role. In his view, one cannot remember that *p* unless it is true that *p*. So he wonders how a competence-based epistemology might countenance false justified beliefs about the past, given (a) that for memory to yield one's belief that *p* is for one to remember that *p*, and (b) that remembering is factive, that one cannot remember what was not so.

Let me first point out how easy it is to have beliefs about the past that do not derive from memory, since one can infer new things about the past. What is more, one can remember things about the present or future: when the next colloquium will take place, for example. So, the issue is not justified belief about the past, but justified belief that derives from memory, and how it could ever be false.

Similar questions can be raised, moreover, about our other faculties. How could an exercise of perception yield false justified belief, for example, if to derive your belief that *p* directly from an exercise of perception is to perceive that *p*? Perceiving, after all, is factive, so one could not possibly perceive that *p* unless it was true that *p*.⁵

All such cases are to be handled in a uniform way, or so I will argue. In each case we have a fallible competence in play, whether it be perception, memory, or perhaps deductive inference. All that is required for any such faculty to count as a faculty, as a competence, is that in the circumstances normal (or better) for its operation, it operate reliably enough: i.e.,

that its yield in such circumstances would turn out true probably enough. This is easily compatible with its occasionally yielding a false belief even in normal circumstances. More easily yet is it compatible with its doing so in *abnormal* circumstances. In all such cases, where it is the faculty or competence itself that is at work, the believer believes competently or with epistemic justification.

Suppose one misremembers a phone number by one digit, where one has no reason to suspect one's normally excellent memory. Here memory operates to yield a false justified belief, although one can hardly remember the phone number to be something that it really isn't, since one gets it wrong by one digit. Still, one is surely better justified in one's belief about the number so long as its content is of a sort that one normally can retain with extremely high reliability, and one has not the slightest reason to mistrust one's memory in this particular case.

Similarly, if one takes oneself to see a red surface, and one has no reason to suspect that the lights are abnormal, then one may be perfectly well justified on a sensory, experiential basis in believing that surface to be red. For, one's belief may then derive from the operation of a reliable color perception that hardly ever leads one astray when exercised in normal conditions, and one has no reason to suspect the relevant conditions holding at the time.

I turn now to the second main problem attributed to my view. The problem, which is said to affect reliabilist views generally, turns on the "Connection Thesis," namely that "if you're justified in believing that there is a reliable information source according to which X, then you are justified in believing X." What, specifically, is the problem?

"JX" is stipulated to mean that the subject is justified in believing X, and "RX" to mean that there is a reliable source of information according to which X. So, we can abbreviate Vogel's Connection Thesis thus:

CT If J(R(X)), then J (X)

Competence-centered virtue epistemologists (like myself), and reliabilists more generally, of whatever stripe, are said to face a serious problem because of CT. Consider a pure reliabilist theory for which

E $J(X)$ is equivalent to $R(X)$.⁶

When joined to CT, this entails

RCT If $R(R(X))$, then $R(X)$.

However, instances of the following schema are not true in general:

If $R(\dots)$, then \dots .

And we have no reason to suppose that RCT in particular would be true when the schema is not true in general. Vogel accordingly doubts the truth of RCT, and takes this to discredit any pure reliabilism committed to E.

This line of objection founders on the fact that reliabilism is a view about doxastic justification, not propositional justification. Vogel himself initially grasps this fact but later it must slip his grip. If we hold on tight to it, then we must make sure to read " $J(X)$ " so that it means that the subject's belief (i.e., his actual believing) of X is doxastically justified. And we must redefine " $R(X)$ " to mean that the subject's actual token belief (believing), $B(X)$, has a source S – a competence or mechanism or process – from which $B(X)$ actually derives, such that S is truth-reliable enough.

Now both CT and RCT seem obviously refuted by the same simple datum: that even if one believes, no matter how justifiably, that $R(X)$ – i.e., that one's belief of X derives from a reliable source it does not follow that one so much as believes X . Much less does it follow that one believes X and does so reliably.

More plausible principles in a reliabilist framework would read rather as follows:

CT If (i) $J(R(X))$, and (ii) $B(X)$ is based on $B(R(X))$, then (iii) $J(X)$.

RCT If (i) $R(R(X))$, and (ii) $B(X)$ is based on $B(R(X))$,
 then (iii) $R(X)$.

These may still turn out to be unacceptable, but they could hardly be refuted by direct appeal to their obvious falsity. For, the basing relation that, according to (ii), links the consequent belief, $B(X)$, with the perspectival antecedent belief, $B(R(X))$, would itself apparently help to constitute a reliable source involved in the truth of the consequent $J(X)$, the reliable source that yields the subject's belief $B(X)$.

True enough, it would not follow apodictically that just because $R(X)$ is believed reliably, that therefore when one bases one's belief of X in part on the reliable belief that $R(X)$ is reliably believed, X itself is reliably believed. For, there might be an instance where a belief of $R(X)$ has a reliable enough source without this yielding a reliable enough source for a belief of X itself. In other words, the truth of one's belief of $R(X)$ might be probabilified by its own source to a level above the relevant threshold, even while one's belief of X was not probabilified above the threshold by being based on belief of $R(X)$. When belief of $R(X)$ derives from a reliable source, then, it does not necessarily follow that a belief of X itself, when based on that fact, would also have a reliable source. Still, it may tend to have a reliable source in general, except for the comparatively small range of cases clustered around the threshold.

In keeping with these most recent reflections, we must reconsider the Connection Thesis. Even if some source for belief of $R(X)$ is reliable above the relevant threshold, basing our belief of X on that very basis may not provide our belief with a source whose reliability is also above the threshold. What seems true is more along the following lines: the more reliable is the source of $B(R(X))$, the more reliable is the source of $B(X)$ that is constituted by its being based on $B(R(X))$. Working our way back to the Connection Thesis, then, its more defensible variant would say rather that the degree of reliability that one can reliably enough attribute to one's overall source for believing X will tend to ensure a similar degree of reliability for one's basing of one's belief of X

on that aforecited attribution. And this fits unproblematically with competence-centered epistemology.

NOTES

¹ These comments are in two parts. The first, drawn directly from the original conference session on *Epistemic Justification* (Blackwell, 2003) at the Inland Pacific Philosophy Conference, focuses on a disagreement concerning internalist foundationalism among three of us involved in that session, namely Larry BonJour, Rich Feldman, and I. In a second section I respond to post-conference comments by Jonathan Vogel directed at my views.

² “[Where]...I have a conscious sensory experience, I am, as already argued, aware of its specific sensory content simply by virtue of being in that experiential state. And therefore if (i) an apperceptive belief that I entertain purports to describe or conceptually characterize that perceptual content, albeit no doubt incompletely, and (ii) I understand the descriptive content of that belief, i.e., understand what an experience would have to be like in order to satisfy that conceptual description, then I seem to be in a good, indeed an ideal, position to judge whether the conceptual description is accurate as far as it goes; and if it is, to be thereby justified in accepting the belief. Here again there is no reason to think that mistake is impossible and thus no reason to think that such an apperceptive belief is itself infallible. But as long as there is no special reason for suspecting that a mistake has occurred, the fact that such a belief seems to accurately characterize the conscious experience that it purports to describe provides, I suggest, an entirely adequate basis for thinking that the description is correct, and thus an adequate basis for justification.” L. BonJour, “Foundationalism and the External World,” *Philosophical Perspectives*, 13, *Epistemology*, 1999; p. 235. Similar statements are found in his “The Dialectic of Foundationalism and Coherentism,” in J. Greco and E. Sosa, eds., *The Blackwell Guide to Epistemology* (Blackwell Publishers, 1999), p. 134; and in his contribution to our current debate, within these covers.

³ Compare footnote 24 of my “Reply to BonJour,” in *Epistemic Justification*, by L. BonJour and E. Sosa (Blackwell, 2003) for an account of such rational justification, and a suggestion that it may be found in the work of Roderick Chisholm. Its correlate for ethics may be seen in Kant’s example of the shopkeeper who is honest exclusively out of self-interest.

⁴ For earlier attempts to do so, see Chapter 7 of *Epistemic Justification* (Blackwell, 2003), prefigured in *Knowledge in Perspective* (CUP, 1991), esp. pp. 84–85; cf. section 5 of John Greco’s Introduction to *Ernest Sosa and His Critics* (Blackwell, 2004).

⁵ The apparently factive character of exercises of our cognitive faculties raises a set of issues discussed extensively in “Knowledge and Intellectual Virtue,” Chapter 13 of my *Knowledge in Perspective* (Cambridge University Press, 1991).

⁶ As John Turri tells me, this needs to be tweaked, since as J and R are defined, one is relativized to a subject, while the other is not. Also please note that I am using expressions in capital letters sometimes as sentences (“B(X),” for example, as the sentence “S believes X”) and sometimes as singular terms (with “B(X)” now representing “S’s belief (instance of believing) of X,” and “J(X)” representing “the proposition that S is justified in believing X”). I trust that context will disambiguate adequately, though sometimes I italicize the expression in question when it is used not as a sentence but as a singular term, though even then the context must still dispose of the remaining ambiguity (between, for example, the token state of believing X, the type state of believing X, and the proposition that one believes X).

Department of Philosophy
Brown University
Providence, RI 02912
USA

Department of Philosophy
Rutgers University
New Brunswick, NJ 08901
E-mail: sosa@brown.edu