

[Published in *American Philosophical Quarterly* 43 (2006): 147–58.  
Official version: <http://www.jstor.org/stable/20010233>.]

## Phenomenal Conservatism and the Internalist Intuition

*ABSTRACT:* Externalist theories of justification create the possibility of cases in which everything appears to one relevantly similar with respect to two propositions, and yet one proposition is justified and the other is not. Internalists find this difficult to accept, because it seems irrational in such a case to affirm one proposition and not the other. The underlying internalist intuition supports a specific internalist theory, Phenomenal Conservatism, on which epistemic justification is conferred by appearances.

### 1. The Internalism/Externalism Divide

Epistemic internalism and externalism may be understood either as views about the nature of knowledge, or as views about the nature of epistemic justification.<sup>1</sup> The focus of this paper will be on epistemic justification. In the following, a particular counter-intuitive consequence of externalist theories of justification is identified, one whose implausibility forms the core of the case for internalism. The argument for internalism is then extended to show that a particular theory of justification, Phenomenal Conservatism, is the preferred form of internalism, in that it is most faithful to the central internalist intuition.

Both internalists and externalists accept a certain supervenience thesis, to the effect that, when one is justified in believing a proposition, there are some conditions *in virtue of which* one is justified in believing it, some conditions that *constitute* or *confer* epistemic justification. Hereafter, a justification-conferring condition will be taken to be a non-redundant part of a sufficient condition on justification,<sup>2</sup> although the arguments following can equally well be stated in terms of either necessary or sufficient conditions on justification. The dispute between internalists and externalists concerns whether these justification-conferring conditions are—in a sense to be defined—entirely internal, or at least partly external. The following five accounts of the distinction invoke different interpretations of “internal” and “external” conditions:<sup>3</sup>

#### *Internal State Account:*

Internalism holds that all of the conditions that confer justification are internal to the subject’s mind (or: supervene on the subject’s mental states).

Externalism holds that at least some of the conditions that confer justification are external to the subject’s mind (or: fail to supervene on the subject’s mental states).

#### *Actual Awareness Account:*

Internalism holds that all of the conditions that confer justification must be conditions that the subject is aware of.

Externalism holds that at least some of the conditions that confer justification may be conditions that the subject is unaware of.

*Potential Awareness Account:*

Internalism holds that all of the conditions that confer justification must be conditions that the subject *can* be aware of.

Externalism holds that at least some of the conditions that confer justification may be conditions that the subject cannot be aware of.

*Introspectability Account:*

Internalism holds that all of the conditions that confer justification must be conditions that the subject can be *introspectively* aware of.

Externalism holds that at least some of the conditions that confer justification may be conditions that the subject cannot be introspectively aware of.

*Appearance Account:*

Internalism holds that all of the conditions that confer justification supervene on how things seem to the subject.

Externalism holds that at least some of the conditions that confer justification fail to supervene on how things seem to the subject.

For reasons discussed below, the Appearance Account offers arguably the most interesting way of drawing the distinction. But there is no need at this point to attempt to adjudicate among these competing accounts of the distinction.

For illustrative purposes, it is worth looking at how each of the above accounts can be used to classify two particular epistemological views. Consider first the following paradigmatic externalist principle of justification:

*Reliabilism:*

S is justified in believing that  $p$ , if S formed the belief that  $p$  by a reliable method, S has no beliefs that either support  $\sim p$  or support the proposition that S formed the belief that  $p$  by an unreliable method, and S has no available reliable belief-forming method that, if used, would have led S to believe that  $\sim p$ .<sup>4</sup>

This principle states a sufficient condition on justification, of which the reliability of S's method of forming the belief that  $p$  is a non-redundant part. The reliability of S's belief-forming method, in turn, is an external condition, in the sense that it does not supervene on the subject's mental states, subjects may be unaware of whether it obtains, subjects may not even be in a position to become aware of whether it obtains, subjects are rarely if ever in a position to become *introspectively* aware of whether it obtains, and it is possible for things to appear the same to the subject in a case in which the condition obtains as they do in a case in which the condition does not obtain. Thus, this view seems to come out externalist on any account of externalism.

The following is a paradigmatic internalist view:

*Phenomenal Conservatism:*

If it seems to S that  $p$ , then, in the absence of defeaters, S thereby has at least some degree of justification for believing that  $p$ .<sup>5</sup>

The main justification-conferring condition in this theory is the condition of its seeming to one that  $p$ . This condition is internal in the sense that it is a mental state of one's own; it is, plausibly, a condition that the subject can become aware of, and introspectively aware at that; and it supervenes on how things appear to the subject. It is unclear whether it satisfies the actual awareness requirement, but if it does not, then the Actual Awareness account of internalism should be rejected as saddling internalists with a needlessly implausible view. This is because if even its seeming to one that  $p$  does not count as an internal condition—perhaps because subjects need not have second-order representations of their own seemings—then there will be precious few, if any, internal conditions for the Actual Awareness Internalist to rely on; it seems equally plausible (or implausible) to hold that subjects may lack awareness of their beliefs, perceptual experiences, and other mental states as that they may lack awareness of how things appear to them.<sup>6</sup>

The no-defeater condition alluded to in the statement of Phenomenal Conservatism is perhaps harder to classify. But if defeaters themselves are characterized in terms of how things appear to the subject, or the subject's beliefs, or something generally along those lines, then the no-defeater condition will be similarly internal. Thus, Phenomenal Conservatism seems to come out as an internalist theory on any plausible account of internalism.

## **2. The Central Internalist Intuition**

Suppose that some condition, E, is external in one of the senses delineated above and is a non-redundant part of some sufficient condition, C, on justification. It seems very likely, then, that it would be possible to construct a case in which, for some person S and some propositions  $p$  and  $q$ ,

- i) S satisfies E with respect to  $p$  but not with respect to  $q$ ;
- ii) S satisfies all the *other* parts of C, with respect to each of  $p$  and  $q$ ;
- iii) S does not satisfy any sufficient condition on justification other than C, with respect to either  $p$  or  $q$ ; and
- iv) It seems to S that he is in the same epistemic position with respect to  $p$  as he is in with respect to  $q$ , and S has no reason for suspecting that either proposition is more justified or more likely to be true than the other.

For example, suppose that Reliabilism is a correct statement of a sufficient condition on justification. Then “the method by which S formed the belief that  $p$  is reliable” is an external condition that is a non-redundant part of a sufficient condition on justification. Now imagine the following scenario.

*The Clairvoyant Brain:*

Susan has two interesting ways of forming beliefs: apparent sensory perception, and clairvoyance. Susan, unfortunately, is a brain in a vat, and her sense perception is entirely unreliable, although she has no reason to suspect this, and it *seems* to her that her sense perception is just as reliable as her clairvoyance. It also seems to her that her perceptual beliefs are adequately justified, just as much as her clairvoyant beliefs. Nor has she any reason to doubt any of this. Oddly enough, Susan actually has psychic powers, and her clairvoyance is highly reliable. One day, Susan seems to see a dog in front of her and has no special reason to doubt the dog's reality. She also has a clairvoyant experience of a purple unicorn grazing in a field somewhere, with no special reason to doubt the unicorn's reality. In fact, there is no relevant dog of the sort she seems to perceive, but there is a real unicorn that she is accurately detecting clairvoyantly.<sup>7</sup>

In this scenario,

- i') Susan has a reliable belief-forming method that tells her that there is a purple unicorn, and (only) an unreliable method that tells her that there is a dog;
- ii') Susan has no beliefs and no other reliable belief-forming methods that would cast doubt on the existence of either the unicorn or the dog, nor on the reliability of either her sense perception or her clairvoyance;
- iii') Susan does not (according to the Reliabilist) have any other source of justification for believing in either the unicorn or the dog, apart from satisfying Reliabilism;<sup>8</sup> and
- iv') It seems to Susan that she has just as much reason for believing in the dog as she has for believing in the unicorn, that the former belief would be just as reliable as the latter, and in general, that she is epistemically no worse positioned with respect to the proposition that there is a dog than she is with respect to the proposition that there is a unicorn. Susan has no reason to doubt any of this, nor any reason to think that it is any more likely that the unicorn exists than that the dog exists.

If Reliabilism is a correct theory, then Susan would be justified in believing that the purple unicorn exists, but unjustified in believing that the dog exists.<sup>9</sup> Internalists will find this counter-intuitive. To see why, consider things from Susan's point of view. It seems that, if one has adequate justification for believing that *p* and none for believing that *q*, it follows that one rationally *should* (or at least *may*) believe that *p* while refraining from believing that *q*. In addition, it seems that, at least normally, including in scenarios like the above, a rational person might recognize and report their own doxastic situation. Thus, Susan might say something like the following to one of the people in her virtual world; call remarks of this form *the Absurd Speech*:

*Absurd Speech:*

I seem to be aware of a dog, just as I seem to be aware of a unicorn. These two experiences seem equally reliable to me, and in general, seem alike in all epistemically relevant respects. However, I believe that there is a unicorn, and I do not believe that there is a dog. I have no reason to think that the unicorn experience is any more likely to be accurate than the dog experience; I just accept the content of the one and not the other, for no apparent reason.

On Reliabilism, the above would be a rational thing for Susan to say—or, more importantly, would (if true) be a report of an epistemically rational state of mind. But the Absurd Speech is not a rational thing to say. Nor is this a matter merely of the propriety of *asserting* what the Absurd Speech asserts; even to think to oneself what the Absurd Speech says would be a mark of irrationality.<sup>10</sup>

The claim that if Reliabilism is true, the Absurd Speech is rational, does not rest on any strong internalist assumptions, such as that individuals are always aware of their own mental states or of what constitutes a reason for what. It requires only that, in a situation of the kind described, Susan might rationally make a correct report of her mental state, including how things seem to her, what she believes and does not believe, and what she has or lacks reason for thinking. Although there exist situations in which a person is not in a position to correctly and justifiably report some of their own mental states, there is no reason for an externalist to deny that Susan could be in a position so to report her mental states in the sort of situation we are imagining; there is no reason, for example, for denying that Susan might have a reliable introspective mechanism.

But perhaps if externalism is true, Susan might be unaware of some of the *reasons* for belief that she has, or might justifiably make false judgements about what reasons she has. For example, Susan might not be in a position to recognize (a) that the actual reliability of her clairvoyance, coupled with the actual unreliability of her sensory experiences, is a reason for thinking that her unicorn experience is more likely to be true than her dog experience, or (b) simply that the occurrence of her unicorn experience is a reason for believing that there is a unicorn, while the occurrence of her dog experience is not a reason for believing that there is a dog. If so, then the Absurd Speech is an incorrect report of Susan's situation; specifically, Susan speaks falsely in saying, "I have no reason to think that the unicorn experience is any more likely to be true than the dog experience." Even so, Susan would be *justified* in saying what she does and justified in her doxastic attitudes towards the dog and the unicorn, and this is counter-intuitive. Moreover, we can modify the Absurd Speech as follows, resulting in remarks that the Reliabilist is committed to regarding as both true and justified:

I seem to be aware of a dog, just as I seem to be aware of a unicorn. These two experiences seem equally reliable to me, and in general, seem alike in all epistemically relevant respects. However, I believe that there is a unicorn, and I do not believe that there is a dog. I don't seem to have any reason to think that the unicorn experience is any more likely to be accurate than the dog experience; I just accept the content of the one and not the other, for no apparent reason.

Again, this does not seem to be a possible accurate report of a rational state of mind.

The preceding argument suggests a natural characterization of the central intuition of internalism about justification. It is that there cannot be a pair of cases in which everything seems to a subject to be the same in all epistemically relevant respects, and yet the subject ought, rationally, to take different doxastic attitudes in the two cases—for instance, in one case to affirm a proposition and in the other to withhold.

What is it for things to seem the same “in all epistemically relevant respects”? In the Clairvoyant Brain case, Susan experiences two appearances: one of a dog, and another of a unicorn. These appearances are not the same in all respects, since, for one thing, they have different contents. But this difference in their contents is not epistemically relevant—that is, it could not plausibly be maintained that Susan is either more or less justified in relying on her perceptual experience than on her clairvoyant experience because the one experience is *of a dog* while the other is *of a unicorn*. On the other hand, such features as the clarity and firmness of the two appearances, how well each coheres with other appearances, and the presence or absence of second-order appearances (such as the appearance that one or another appearance is reliable), are plausibly taken to be epistemically relevant. In the Clairvoyant Brain case, by stipulation, those sorts of factors are comparable for the two appearances—that is, the dog and unicorn appearances are equally clear and firm for Susan, they cohere equally well with other appearances, they seem to Susan to be equally reliable, and so on. It is in that sense that things seem the same to Susan, with respect to the proposition that there is a dog and the proposition that there is a unicorn, in all epistemically relevant respects. The reliabilist nevertheless rules the one proposition justified and the other unjustified.

All externalist theories of justification apparently permit the construction of cases of this kind. To see why, consider the characterizations of the internalism/externalism dispute from section 1 above. On the Internal State Account, the externalist will presumably affirm that there could be two cases in which a subject’s mental states are relevantly the same, but the subject has justification for belief in one case and not in the other, because of a difference in some factor external to the mind. On the Actual Awareness (Potential Awareness, Introspectability) Account, the externalist will affirm that there could be two cases in which everything the subject is aware of (can be aware of, can be introspectively aware of) is relevantly the same, but the subject has justification in one case and not in the other, because of a difference in some factor the subject is unaware of (cannot be aware of, cannot be introspectively aware of). In each of these cases, it seems that the external factor in question would not, or at least need not, affect how anything appears to the subject. Of course, the point is clearest on the Appearance Account of the internalism/externalism distinction; on this account, externalism is essentially *defined* as a view according to which it is possible to have pairs of cases of the problematic kind, in which all appearances are relevantly the same but justification is present in the one case and not the other.<sup>11</sup> The Appearance Account is perhaps the most illuminating because it so

directly connects the definition of “internalism” with the central intuition behind internalism.

### 3. An Extension of the Argument: The Acquaintance Theory

The argument of section 2 can be extended to some theories that might be considered internalist. The Acquaintance Theory of Non-Inferential Justification is one such theory. A simple version of the theory is the following.

*Simple Acquaintance Theory:*

S has non-inferential justification for believing that  $p$ , if and only if S is acquainted with the fact that  $p$ .

Acquaintance is taken to be a *sui generis* relation that one may stand in to a fact, a kind of direct awareness of the fact. Note that one’s being acquainted with the fact that  $p$  entails that  $p$  is true (“acquaintance” is factive). Typically, advocates of acquaintance theories take the possible objects of acquaintance to include (facts about) one’s own mental states and certain abstract objects.<sup>12</sup>

One problem for the Simple Acquaintance Theory is the problem of fallibility: we are fallible in all our major kinds of beliefs, including introspective and intuitive beliefs. Thus, suppose that you are simultaneously acquainted with two distinct pains. It seems to you that they are equally intense, and you have no reason for doubting this. However, one of the pains is in fact slightly less intense than the other. (The fallibility of your introspective sense is particularly understandable if the pains are of different kinds, though it is possible even if they are pains of the same kind.<sup>13</sup>) The next day, you are in a similar situation, except that this time, the two pains are actually equally intense, and you are directly acquainted with this fact. Presumably, this—the equal intensity of two present, conscious pains—is the sort of thing with which one might be acquainted, according to acquaintance theorists. In neither case do you have any defeaters for the relevant proposition, nor have you any ulterior justification for either proposition.

According to the Simple Acquaintance Theory, you lack justification for believing that the pains are equally intense on the first occasion, since you lack genuine acquaintance with their being equally intense, but you have justification for believing the pains are equally intense in the second case. So you should withhold judgement in the first case and believe in the second case. You do so, and you retain the relevant states of withholding and believing after all the pains are over. You then say:

I think that the second pair of pains were equally intense, but I don’t think the first pair were. The first pair seemed to me to be equally intense, just as the second pair did, and I have no reason for doubting this, nor any reason to think the second introspection would be more reliable than the first. I just believe in the one case and not in the other, for no apparent reason.

Some readers may dispute the possibility of the case or my description of it. But it seems very likely that there are at least some possible cases in which a subject appears to be introspectively acquainted with a fact, lacks grounds for questioning his introspection, and yet is not actually acquainted with a relevant fact. This is all that the argument requires. In any such case, the Simple Acquaintance Theory would counsel one to eschew the relevant introspective belief, although said belief would appear to oneself to be relevantly similar to normal introspective beliefs that the Simple Acquaintance Theory approves of.

Doubtless many will find the Simple Acquaintance Theory too simple in any case. It is obviously ill-suited from the start to deal with the possibility of justified but mistaken intuitions and introspections, which most epistemologists would wish to recognize. How, if at all, might an acquaintance-based theory of non-inferential justification be modified to allow for this possibility? Here is one suggestion, derived from Richard Fumerton:

*Acquaintance/Similarity Theory:*

S has non-inferential justification for believing that  $p$ , if and only if S is acquainted with a fact that is sufficiently similar to the fact that  $p$ .<sup>14</sup>

This view naturally raises questions about what will constitute sufficient similarity. But there is some initial plausibility to the thesis. It seems to explain the intuition that, in the pain cases described above, you were justified both times in believing that the pains were equally intense. What you were acquainted with in the first case might actually have been the fact that the two pains were *almost* equally intense, which is very similar to their being equally intense. So you were justified in believing that the two pains were equally intense.

My case against the Acquaintance/Similarity Theory turns on the possibility of cases that *seem* to the subject just like cases of acquaintance with a fact that  $p$  but that do not involve acquaintance with any fact that is genuinely similar to  $p$ . Consider the following example.

*Indignation/Resentment Case:*

Assume that indignation and resentment are distinct emotions, which people sometimes mistake for one another. Sam has a feeling of resentment, which falsely appears to him to be one of indignation. Imagine further that, although indignation and resentment *seem* very similar to Sam, they actually have a number of theoretically important dissimilarities; for instance, the capacities for the two emotions evolved independently, at different times in the history of our species; they serve different evolutionary functions; they have importantly different causes in the environment; they are proximately caused by (or realized in) different brain states; and they tend to lead to different behavior. For all of these reasons, in the completed scientific theory of the mind, they will be classified as very different emotions. Nevertheless, again, Sam finds it very difficult to distinguish them introspectively.



The Acquaintance/Similarity Theory implies that Sam is not justified in believing that he feels indignation. Now suppose that the next day, Sam experiences a genuine case of indignation, with which he is acquainted, and which seems to him indistinguishable from his first experience, which was actually one of resentment. On the Acquaintance/Similarity Theory, Sam can rationally say, “Today, as yesterday, I seem to be experiencing indignation. Today’s emotion *seems* to me just like yesterday’s. But while I believe that I am experiencing indignation today, I do not believe I was experiencing indignation yesterday, although I have no reason to think my introspection today would be any more reliable than yesterday.”

We might avoid this scenario by replacing the Acquaintance/Similarity Theory with an Acquaintance/*Apparent* Similarity Theory:

*Acquaintance/Apparent Similarity Theory:*

S has non-inferential justification for believing that *p*, if and only if S is acquainted with a fact that appears to S to be similar to the fact that *p*.

This seems to better capture what we were getting at with the Acquaintance/Similarity Theory. It is the fact that indignation and resentment *seem similar to Sam*—rather than that they are similar in an objective sense—that explains why his belief that he is feeling indignation is reasonable. But the Acquaintance/*Apparent* Similarity Theory is surely false; if S is acquainted with a fact that seems to him merely to be *similar* to the fact that *p*, but does not seem to actually *be* the fact that *p*, then presumably S would be irrational to believe that *p* on that basis. For example, suppose that on introspecting, Sam finds that he seems to be feeling resentment, and suppose that it also seems to him that this resentment is quite *similar* to indignation, though it does not seem to actually *be* indignation. Then Sam would be irrational to conclude that he is feeling indignation.

In short, fallibilistic acquaintance-based theories of non-inferential justification run into problems analogous to those facing epistemic externalism. Once one grants that one can be non-culpably mistaken about whether one is acquainted with the fact that *p*, the way is open for cases that are, to the subject, indistinguishable from genuine cases of acquaintance but are not in fact cases of acquaintance. An acquaintance-based theory of non-inferential justification will therefore rule it reasonable in some cases to accept one proposition while withholding or denying another, even though one can identify no relevant difference between the two propositions.

#### **4. Phenomenal Conservatism as the Preferred Form of Internalism**

Following the characterizations given in section 1 above, the following versions of epistemic internalism may be formulated.

*Internal State Internalism:* All of the conditions that confer justification supervene on the subject’s mental states.

*Actual Awareness Internalism:* All of the conditions that confer justification must be conditions that the subject is aware of.

*Potential Awareness Internalism:* All of the conditions that confer justification must be conditions that the subject can be aware of.

*Introspectability Internalism:* All of the conditions that confer justification must be conditions that the subject can be introspectively aware of.

*Appearance Internalism:* All of the conditions that confer justification supervene on how things seem to the subject.

More than one of these versions of internalism may be true. Indeed, it is plausible that all of them save the Actual Awareness version are correct. Nevertheless, there is a point to asking which is the best formulation of internalism, in the sense of the formulation that best reflects the underlying motivation for epistemic internalism. There is reason to believe that the answer is the Appearance formulation.

The sort of intuitions invoked in sections 2 and 3 above seem to comprise the main underlying motivation for internalism. Externalist theories of justification countenance peculiar cases in which subjects rationally believe (disbelieve, withhold) some proposition yet are seemingly in no position to account for why they should believe (disbelieve, withhold) as they do. This is what internalists object to. A rational person, it seems, ought to be in a position, if he should come to reflect on his doxastic attitudes (that is, his attitudes of belief, disbelief, or suspended judgement toward various propositions), to approve those attitudes as justified—or at least he should not be in a position such that on reflection, he would or should *disapprove* of those attitudes. Such cases as that of the Clairvoyant Brain bring out externalism's violation of this constraint. In the Clairvoyant Brain Case, it seems that Susan (the brain) is in no position to reflectively endorse both her credulity towards her clairvoyant experiences and her skepticism towards her sensory experiences; from her own point of view, the combination of attitudes must appear arbitrary and inexplicable. It therefore seems irrational of her to persist in those attitudes; yet that is what the externalist counsels.

Consider, now, how well each of the above versions of internalism reflects this core internalist intuition. Internal State Internalism fares poorly. It is not that Internal State Internalism is false—indeed, it is entailed by Appearance Internalism, given that appearances are mental states; but Internal State Internalism is not specific enough to explain why the Absurd Speech is absurd. Justification-conferring conditions do not merely supervene on subjects' mental states; they supervene, more specifically, on how things appear. This is brought out by cases such as the Indignation/Resentment Case. There, we saw that what mattered was not so much what mental state Sam was in, as what mental state he appeared to himself to be in. When Sam shifted from feeling resentment on one day to feeling indignation on the next, this made no difference to what he was justified in believing, *because things still appeared to him the same*. An Internal State Internalism that does not embrace Appearance Internalism will mishandle

this sort of case—if, that is, we say that justification depends on mental states that do *not* supervene on how things seem to the subject, then we will be able to construct a pair of cases in which the subject has relevantly different mental states but yet things appear the same to the subject—and we will be forced to say that the subject ought to take different doxastic attitudes in the two cases. The subject will then be in a position to make a version of the Absurd Speech. This is just the sort of would-be possibility illustrated by the Indignation/Resentment Case.

Once we accept this, we can see that similar points apply to Actual Awareness, Potential Awareness, and Introspectability Internalism. Assuming that “awareness” is a success term, a change in what a subject is or can be *aware* of can be effected purely by a change in the actual facts, without any change in how things appear to the subject. This is illustrated, again, by the Indignation/Resentment Case, in which there is a difference in what the subject is and can be aware of, and specifically in what the subject can be introspectively aware of—on the first day, Sam lacks (actual or potential) introspective awareness of indignation, because the emotion he is feeling is not in fact indignation; but on the second day, he has introspective awareness of indignation. Yet this difference in what Sam is and can be introspectively aware of seems to make no difference to what Sam is justified in believing, because things *appear to Sam* the same.

In short, if we deny Appearance Internalism, then even if we embrace some other version of internalism, we will have to allow the possibility of pairs of cases in which everything seems to a subject the same in relevant respects but yet the subject is justified in believing in the one case and withholding or disbelieving in the other. Thus, the form of internalism supported by the argument of sections 2 and 3 above is Appearance Internalism.

Now, what is the relationship between Appearance Internalism and Phenomenal Conservatism? Appearance Internalism articulates a constraint on justification-conferring conditions, and Phenomenal Conservatism identifies a particular justification-conferring condition that satisfies that constraint. Moreover, Phenomenal Conservatism is the most natural and plausible way to articulate such a condition. For, if appearances determine what we are justified in believing, then the most natural account of how they do that is that the appearance that *p* confers justification for believing that *p*—at least to some degree, and at least in the absence of defeaters. This does not rule out that appearances may also bear on justification in other ways—for instance, it does not rule out the principle that the appearance that a belief was formed in an unreliable way may defeat that belief’s justification. But it is very difficult to think of a plausible and natural epistemological principle that permits some beliefs to be justified in virtue of how things seem to the subject, without invoking the idea that the appearance that something is the case itself provides some sort of justification for believing that very thing. If a subject is to be justified in believing that *p* in virtue of the way things appear, what aspect of the appearances could be more relevant than the appearance that *p* is the case? And insofar as that is relevant, surely it is positively, rather than negatively relevant—surely, that is, we want to say that the appearance that *p* is the case contributes to *p*’s justification, as opposed to detracting from it. Other plausible connections between appearances and justification seem to be parasitic on Phenomenal Conservatism; thus, consider

again the view that a belief's appearing to be formed in an unreliable way defeats its justification: it seems that this would only be true if the belief's appearing to be formed in an unreliable way provided some degree of justification for thinking that it *was* formed in an unreliable way, which is a special case of Phenomenal Conservatism.

Something like Phenomenal Conservatism seems to underlie our intuitive reactions to the scenarios discussed above and to the various Absurd Speeches that those scenarios would lead to on externalist theories of justification. For example, in the Clairvoyant Brain Case, Susan's combination of attitudes towards her dog experience and towards her unicorn experience seems irrational, because the dog seems to her just as real as the unicorn, the two experiences seem equally likely to be veridical, and she has no apparent reason for doubting either. All of this would render her combination of attitudes irrational only if these seemings were relevant to what Susan was justified in believing in roughly the way that Phenomenal Conservatism indicates—that is, if the dog's seeming equally real as the unicorn conferred at least some sort of justification (in the absence of defeaters) for thinking that it *was* equally real, if the two experiences' appearing equally likely to be veridical was some sort of *prima facie* justification for thinking that they were equally likely to be veridical, and so on.

It may seem that Phenomenal Conservatism is overly permissive, granting justification to a belief whenever the believer *takes* it to be justified, reliable, and so on. There are two reasons why this is not the case. First, what one believes to be the case need not be the same as what appears to one to be the case. In particular, cases of self-deception and accepting a proposition on faith are both plausibly regarded as cases in which one believes what does not appear to be the case; Phenomenal Conservatism thus is not committed to granting even *prima facie* justification in such cases. Second, Phenomenal Conservatism allows for a given belief's relationship to other beliefs and appearances to impact its justification, whether positively or negatively. Thus, for example, a belief's failure to cohere with the rest of a subject's appearances and beliefs might deprive it of justification, even when the belief in question appears to the subject to be true. In such a case, the believer's merely *thinking* that the belief that P coheres with the rest of his system of beliefs and appearances would not prevent his justification for P from being defeated, provided that the principle governing when a belief's justification is defeated adverts to *actual* coherence relations (or lack thereof).

In short, it appears that the central intuitions supporting epistemic internalism support Phenomenal Conservatism in particular. In that sense, Phenomenal Conservatism is the paradigmatic form of internalism, and the form that externalists ought first and foremost to confront.<sup>15</sup>

## References

- Alston, William P. 1988. "An Internalist Externalism," *Synthese* 74: 265-83.  
BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

- . 2000. “A Version of Internalist Foundationalism,” pp. 3-96 in Ernest Sosa and Laurence BonJour, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Malden, Mass.: Blackwell.
- Bergmann, Michael. 1997. “Internalism, Externalism and the No-Defeater Condition,” *Synthese* 110: 399-417.
- Fumerton, Richard. 1988. “The Internalism/Externalism Controversy,” *Philosophical Perspectives 2: Epistemology*, pp. 443-59. Cambridge, Mass.: Blackwell.
- Fumerton, Richard. 1995. *Metaepistemology and Skepticism*. Lanham, Md.: Rowman & Littlefield.
- Goldman, Alvin I. 1992. *Liasons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, Mass.: MIT Press.
- Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham, Md.: Rowman & Littlefield.
- . Forthcoming. “Compassionate Phenomenal Conservatism,” *Philosophy and Phenomenological Research*.
- Kim, Kihyeon. 1993. “Internalism and Externalism in Epistemology,” *American Philosophical Quarterly* 30: 303-16.
- Mackie, J.L. 1980. *The Cement of the Universe*. Oxford: Clarendon Press.
- Russell, Bertrand. 1997. *The Problems of Philosophy*. New York: Oxford University Press.
- Sosa, Ernest. 2000. “Reply to BonJour,” pp. 201-28 in Ernest Sosa and Laurence BonJour, *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Malden, Mass.: Blackwell.

## Notes

<sup>1</sup>See Fumerton (1988), Kim (1993) and Bergmann (1997) on the internalism/externalism controversy. Bergmann argues for construing the internalism/externalism debate as one concerning the nature of warrant, rather than justification.

<sup>2</sup>This notion derives from Mackie (1980, pp. 61-2), who uses it to analyze causation.

<sup>3</sup>The first three of these accounts are based on Fumerton’s (1995, pp. 60-66) characterizations of, respectively, internal state internalism, strong access internalism, and weak access internalism. BonJour (1985, p. 31) seems to endorse an internalist view along the lines of the Actual Awareness Account. Kim (1993, p. 305) is among the many who endorse the Introspectability Account.

<sup>4</sup>This is a simplified version of Goldman’s (1992, pp. 123, 130) view.

<sup>5</sup>See Huemer (2001, pp. 98-115) and (forthcoming) for discussion of this principle, where “it seems to S that *p*” and “it appears to S that *p*” are taken to report a distinct kind of propositional attitude, different from belief, which may be

---

termed an “appearance,” and which often causes one to adopt beliefs. Sensory experience, memory experience, introspective appearances, and intuitions are all species of appearances.

The present formulation of Phenomenal Conservatism differs from its original version in three respects. First, the phrase “at least some degree of” has been added to make clear that one need not have full justification for belief merely by having, for example, a weak and wavering appearance. Second, Phenomenal Conservatism is now treated as a principle governing justification in general, rather than specifically non-inferential justification. Thanks to Michael Tooley, who endorses neither version of the principle, for pointing out the need for such revisions.

<sup>6</sup>In addition, as Fumerton (1995, p. 80-82) and Alston (1988, p. 271) observe, the actual awareness requirement generates an infinite regress. Briefly, assuming that actual awareness of a condition entails justified belief that the condition obtains, the actual awareness account implies that, if one justifiably believe that  $p$ , one must also have a justified belief that some condition,  $C(p)$ , on one’s being justified in believing  $p$  obtains; but then one must also have a justified belief that some condition,  $C(C(p))$ , on one’s being justified in believing that  $C(p)$  obtains, obtains; and so on.

7

Nothing here turns on the apparent silliness of purple unicorns; to Susan, purple unicorns seem no sillier than dogs, nor has she any more reason to be suspicious of purple unicorns than of dogs.

<sup>8</sup>Since Reliabilism as defined above only provides a sufficient condition on justification, a reliabilist could also embrace some other sufficient condition—perhaps even Phenomenal Conservatism. However, unless Reliabilism is at least sometimes the only sufficient condition on justification that a belief satisfies, there would be little point to holding the view—it would be a sufficient condition on justification that is never required to explain why any belief is justified. It is therefore fair to assume that it is possible to construct a case in which the relevant substitution instance of (iii) holds.

<sup>9</sup>Goldman (1992, p. 134) seems to confirm this assessment, declaring unjustified the beliefs of victims of Cartesian demons. Though Goldman allows another, “weak” sense of justification in which such beliefs would be justified, it is his strong sense of justification that is pertinent here. His weak sense appears to be an internalist notion of justification.

<sup>10</sup>The Clairvoyant Brain case may provide an objection to certain forms of internalism as well. BonJour (1985, pp. 41-5; 2000, pp. 28-9, 31-2) discusses a case in which an individual has reliable clairvoyant beliefs with no grounds either for believing or for disbelieving that they are reliable. BonJour finds it intuitive that the clairvoyant beliefs in his case are unjustified. If at the same time, one

---

affirms that perceptual beliefs are justified in the absence of grounds either for affirming or for denying their reliability, then one will be led to endorse an Absurd Speech in the Clairvoyant Brain case. For this reason, the internalist should not embrace this combination of views. Cf. Sosa's (2000, p. 223) response to BonJour.

The Phenomenal Conservative will hold that both clairvoyant beliefs and perceptual beliefs, if based upon appropriate appearances, are *prima facie* justified. This is consistent with the admission that the background knowledge of typical actual humans largely defeats the justification that one would have for clairvoyant beliefs.

<sup>11</sup>Strictly speaking, externalism does not entail that it is possible for a *single subject* to experience two such cases; it entails only that two possible subject-proposition pairs,  $\langle S, p \rangle$  and  $\langle T, q \rangle$ , might be such that  $S$  is justified in believing  $p$ ,  $T$  is not justified in believing  $q$ , and everything appears relevantly the same to  $S$  and  $T$ . But virtually any view that allows this will also allow for the possibility of such cases in which  $S=T$ .

<sup>12</sup>Russell (1997) and Fumerton (1995, pp. 73-9) defend versions of the Acquaintance Theory. See Russell (1997, chapters 5, 10) on the possible objects of acquaintance.

<sup>13</sup>See Williamson 2000, chapter 4, for a persuasive argument that people are fallible in their judgements about all types of mental states.

<sup>14</sup>See Fumerton (1995, p. 77). The right hand side of the biconditional is not to be read as implying that the fact that  $p$  exists. There might be a fact "similar to the fact that  $p$ " even if  $p$  is false, just as there might be a "godlike being" even if there is no God. Though the biconditional could be rephrased in terms of similarity of *propositions* rather than of (would-be) facts, Fumerton's actual view does not recognize the existence of false propositions (1995, pp. 73-4). One could also rephrase the view to speak of states of affairs, perhaps including merely possible states; the arguments following in the text would apply equally well to such a view.

<sup>15</sup>I would like to thank the anonymous referees at *American Philosophical Quarterly* for helpful comments on this paper.