

*This is the final manuscript draft. The published article is available at  
<https://link.springer.com/article/10.1007%2Fs10670-017-9956-5>  
Please cite published version.*

## **AGENT-AWARENESS IN REFLECTIVE KNOWLEDGE**

Sharon E. Mason  
University of Central Arkansas

### ABSTRACT:

I argue that current discussions of the epistemological significance of reflection have entangled concerns about reflection with agential concerns. I begin by showing that a central strand of internalist criticism finds externalism unsatisfactory because it fails to provide a particular kind of *self-knowledge*, knowledge about the epistemic status of *one's own beliefs*. Identifying this internalist motivation as the desire for a kind of self-knowledge opens up new possibilities and suggests new conceptual resources. I employ one of these resources—Richard Moran's distinction between the theoretical stance and the deliberative stance—to locate two types of reflection: *mere reflective awareness* of one's attitudes and *agent-awareness* of one's attitudes. I then examine Ernest Sosa's account of the importance of reflection, showing how Moran's distinction brings out the centrality of agential concerns in Sosa's argument for reflective knowledge. I also consider briefly its relevance to fully apt knowledge. While I focus on Sosa's epistemology, the point extends to internalism more generally.

**AGENT-AWARENESS IN REFLECTIVE KNOWLEDGE**

Of what epistemological significance is the ability to assess one's own epistemic standing? This question gets at key issue in a well-worn epistemological controversy over the last several decades—the debate between internalism and externalism. I argue that the idea that reflection is epistemologically significant is often confused with other concerns that are more properly understood as agential. In this paper, I focus on Ernest Sosa's defense of *reflective knowledge* and his development of *fully apt knowledge*. While Sosa is no internalist, he has offered an extensive defense of reflection, and his defense focuses in particular on the importance of knowledge that is of a better "kind" than mere externalist knowledge. Furthermore, as Sosa's view has developed, agency has played an increasingly central role in motivating the higher levels of knowledge that he proposes. However, the connection between reflection and agency is not readily apparent. I propose that we need to think more carefully about the way in which knowledge about the epistemic credentials of one's own belief is a kind of *self-knowledge*. Doing so reveals distinct characterizations of reflection that can then be employed to assist in the task of disentangling the reflective from the agential, and ultimately, putting them back together again properly. The disentangling is the focus of this paper, although at the end I will offer a few suggestions for what I think is the right direction of how they may be put back together.

I begin in section 1 by showing that the ability to make an epistemic self-assessment is a primary concern in an influential strand of arguments for internalism. This concern is sometimes obscured in discussions of internalism that focus on the nature of justification or on the priority of various epistemological projects. These self-assessments are instances of self-knowledge about the epistemic status of one's belief. In section 2, I introduce a distinction that Richard

Moran draws between two stances one might take toward one's own attitudes in order to characterize different ways in which one might be aware of one's own attitudes. My goal here will be to address how best to understand the kind of epistemic self-assessments that many internalists value, given that these assessments are instances of self-knowledge. In section 3, I examine Sosa's account of the importance of reflective knowledge, as opposed to mere animal knowledge. Sosa identifies increased *coherence* and *defensibility* as two goods that reflective knowledge affords, but mere animal knowledge does not. I argue that there are two versions of each of these goods: a weak version, and a robust version. If Sosa is interested in the weak version, then it is doubtful that reflection is required to obtain this good. However, if he is interested in the strong version, then reflection alone is insufficient: agent-awareness is required. In section 4, I consider Sosa's most recent defense of the value of reflective knowledge as deriving from fully apt knowledge, and I show how my account offers a friendly amendment to his.

## 1.

I would like to begin by saying something about the way I understand the distinction between internalism and externalism, as these labels have been used to apply to a variety of philosophical positions. Here I will rely on Michael Bergmann's characterization of the debate. While Bergmann describes a variety of ways to carve up the internalist/externalist territory, the view he defends as the most useful way of drawing the distinction is, in a nutshell, that "internalists think there is an awareness requirement on knowledge and externalists think there isn't" (2006, p. 7). I articulate this "awareness requirement" in the following terms: a theory of knowledge is *externalist* if the ability to make an assessment of one's own epistemic standing is not required for knowledge. A theory of knowledge is *internalist* if the ability to make some kind

of assessment of one's own epistemic standing is taken to be a necessary condition for knowledge.

There are, of course, also a variety of ways to understand what it is to assess one's own epistemic standing. A self-assessment may be directed toward one's own epistemic status, so one might say "I am justified in believing  $p$ ", or "I have good evidence for  $p$ ", or "I have good reason to believe that  $p$ ." It may also focus on the epistemic status of one's belief, such that one might say "My belief is justified/reasonable/etc." What is crucial in either case is that the subject recognizes that the belief being assessed is her own, as the locutions "I" and "My" indicate. Disputes over the proper characterization of those epistemic self-assessments (evidence, justification, reasonability, aptness, etc.) are set aside for the present purpose. I am interested in the (internalist) idea that some kind of (actual or possible) assessment of oneself—in particular, one's own epistemic standing—is necessary for knowledge. In my estimation, the most plausible internalist objections to externalism regard epistemic self-assessments as a necessary condition for knowledge. I will focus my attention there.

Here are a few examples of what I have in mind. First, we can see the concern for an epistemic self-assessment in two classic objections to externalism, one developed by Laurence Bonjour and the other by Keith Lehrer. Their scenarios have a similar structure: Bonjour gives us Norman, a clairvoyant whose reliable clairvoyant capacity provides him with true beliefs about the location of the president. In Lehrer's scenario, Mr. Truetemp has a reliable device implanted in his head that produces beliefs about the temperature. Crucially, in both cases, the subjects (Norman and Mr. Truetemp) are unaware of their reliable capacities. Furthermore, in both cases this lack of reflective self-assessment is given as the primary reason for objecting to externalism. Bonjour summarizes the objection in this way:

The intuitive difficulty with externalism that the following discussion is intended to delineate and develop is this: on the externalist view, a person may be ever so irrational and irresponsible in accepting a belief, when judged *in light of his own subjective conception of the situation*, and may still turn out to be epistemically justified (1980, p. 59, emphasis mine).

In a similar spirit, Lehrer presents his *Opacity Objection*:

There is, however, a general objection to all externalist theories that is as simple to state as it is fundamental: the external relationship might be opaque to the subject, who has no idea that her beliefs are produced, caused, or causally sustained by a reliable belief-forming process or properly functioning cognitive faculty... All externalist theories share a common defect, to wit, that they provide accounts of the possession of information, which may be opaque to the subject, rather than of the attainment of transparent knowledge. (1990, p. 185)

These sorts of cases and the counterexamples to externalism they purport to provide are versions of what Bergmann calls the *Subject's Perspective Objection*:

Subject's Perspective Objection (SPO):

If the subject holding a belief isn't aware of what that belief has going for it, then she isn't aware of how its status is any different from a stray hunch or an arbitrary conviction. From that we may conclude that from her perspective it is an accident that her belief is true. And that implies that it isn't a justified belief. (2006, p. 12)

The SPO concerns the requirements for having a justified belief; however, it is but a short step to knowledge. We only need to add a single conditional: If a subject doesn't have a justified belief, then the subject does not have knowledge. But my concern is not whether the SPO tells us anything about the nature of knowledge. Notice that the objection expresses the idea that knowledge requires the ability to make an assessment of *one's own belief*. The question isn't merely what a particular belief "has going for it" in order to distinguish it from mere guesses (although there certainly is an anti-luck concern operative here: one must be able to tell that one's belief is better than a belief that has perhaps arisen out of the blue). The kind of knowledge that is required here is knowledge about oneself, for it is about one's own mental attitude of belief.

The ability to assess *oneself* epistemically also comes out clearly in a related discussion, where responding to skepticism is identified as the core concern of a particular version of internalism about knowledge (the version of internalism defended by Bonjour and Lehrer). Brett Coppenger, for instance, calls this version of internalism “traditional internalism” and defines it in this way:

...traditional internalism emphasizes rationality’s demand for philosophical assurance, on the basis of evidence that can withstand the strongest skeptical challenges, that our ordinary beliefs (perceptual and otherwise) are true. According to the traditional internalist, the philosopher, qua philosopher, ought to begin the epistemological project from the inside, placing a premium on satisfying our philosophical curiosity. (2016, pp. 1-2)

Coppenger highlights here what Richard Fumerton calls the *fundamental internalist concern*:

...that having knowledge or justification in the externalist sense doesn’t seem to satisfy philosophical curiosity. It doesn’t seem to provide any assurance of the sort the philosopher seeks when wondering about the truth of various propositions. (Fumerton 2004, quoted in Coppenger 2016, p. 4)

The aim of satisfying one’s own philosophical curiosity is about what is true, but it is also about what one should believe as a result of one’s judgments about what is true. Bonjour makes this clear when he observes that the aim of satisfying one’s philosophical curiosity arises only from within one’s own first-person perspective. According to Bonjour, this aim plays a crucial role in motivating the internalist project: “In fact, the central rationale for internalism...arises when I ask not the third-person question of whether someone’s beliefs are true or reliably arrived at, but instead the first-person question about the truth (or reliability) of my own beliefs” (2001, p. 53).

## 2.

So far, I have identified epistemic self-assessments as a primary motivation for the view that externalism about knowledge is problematic. We need not take a position on matters such as whether externalism is not a satisfactory theory of knowledge, as opposed to thinking that

externalism is insufficient to address a particular epistemological project, or that it is insufficient to account for a higher “kind” of knowledge. Despite the disagreement about these issues that we might describe as the *scope* of the problem, there appears to be agreement among internalists about the *nature* of the problem, that is to say, exactly what it is that externalist knowledge lacks. Externalist knowledge lacks an *awareness of the epistemic credentials of one’s beliefs*. Put this way, it might seem that what is lacking is a kind of self-knowledge *about* one’s belief. This is the idea to be explored in detail in this section.

When I know something about the epistemic credentials of a particular belief of mine, I have a kind of self-knowledge. While the point may seem obvious to some, it may be helpful to pause here to examine this idea. When I am aware, for instance, that my belief *p* is aptly formed—according to Sosa, that it is true and that it was formed as the result of an exercise of my competency in suitable circumstances for its use—then I know something about myself; namely, *that I believe that p*. I also am aware that *p* is apt. But I am not merely aware that *someone* has an apt belief; I am aware that this *someone* is *me*. There is parallel here with a famous example in which John Perry considers a shopper who was aware that *someone* was spilling the sugar (he saw the trail of sugar on the floor), but did not yet realize that *he* (the shopper) was the one spilling it. When the shopper thought “Someone is spilling the sugar”, in a certain sense the shopper knew something about himself, for he was the one doing the spilling. However, the shopper lacked the kind of self-knowledge that can be expressed by statements that use the first-personal pronoun “I”. That is to say, although the shopper was in a position to say that someone was spilling the sugar, he was not in a position to say “I am spilling the sugar”.

So too, someone might be in a position similar to Perry’s shopper with regard to the status of the epistemic state of one of his beliefs. For instance, I could be aware that *someone* has

an apt belief, and it might also be the case that this someone is me, without being aware that *my* belief is apt. We need not go through any contortions to come up with a plausible case. Here's one: Suppose I am a participant in a study of the effects of certain kinds of manipulation on a person's beliefs. One day while at a coffee shop, I notice that one of the researchers is sitting next to me. Perhaps I'm bored and feeling a bit nosy, so I glance over at some of the researcher's notes while she is refilling her coffee. I read that in the study, Participants A-F were manipulated while Participants G-L were not. I do not know my own letter; although I am subject L, I am unaware of that fact. In this case, I know that Subject L has an apt belief, but I don't know that *I* have an apt belief.

This analogy helps to illuminate that a self-assessment of one's epistemic state requires a certain kind of self-knowledge. We must be careful here; it's not the case that all the information relevant for making the self-assessment must be available to a person via introspection alone. That is a further requirement that some internalists have defended, but those issues are distinct from what I have in mind. My point is simply that a self-assessment of one's epistemic state requires having one kind of self-knowledge such that one is aware at least that the assessment is of *one's own* belief. Consider again the idea that externalist knowledge is problematic (in whatever sense) because the subject lacks knowledge *about* the epistemic credentials of her belief. We are now in a position to add that the subject lacks a kind of self-knowledge, of being able to recognize that the belief in question is one's own. This self-knowledge includes identifying that one has a particular mental attitude; i.e., *that* one believes that *p*, as knowing that one believes that *p* is required for being aware that one's belief that *p* has something going for it, epistemically speaking.

Now that we are in the territory of thinking about different *kinds* of self-knowledge, it becomes apparent rather quickly that this particular line of thought in traditional internalism intersects with a central issue in the study of the nature of self-knowledge regarding various types of knowledge one might have of one's own mental attitudes. The relevant attitudes are, in this case, the beliefs required to satisfy this concern of the traditional internalist. Again, the concern is often identified as a concern that a knower should be aware of certain facts *about* the epistemic status of her own beliefs. However, as a kind of self-knowledge, knowledge of which attitudes one has is arguably neither the most interesting nor paradigmatic kind of self-knowledge.

Richard Moran's work helps illuminate the difference between being aware of one's attitude in the sense that one *knows the facts about what one's attitudes are*, versus being aware of one's attitude in a sense that Moran identifies as in some sense *agential*. Moran, picking up on a theme in Sartre's work, identifies two stances that it is possible to take toward one's own attitudes. The first is a theoretical stance, in which one is aware of what mental attitudes one has, but, crucially, takes them to be so much information about oneself, in a similar way that one knows one's height, age, eye color, etc.<sup>i</sup> The *theoretical stance* is contrasted with what Moran calls the *deliberative stance*. It will be easier here if we limit our discussion to a particular mental attitude, such as a belief. When one occupies the deliberative stance toward a particular belief, one is able to consider the things one believes in such a way that one treats one's deliberative activity as relevant to the formation and perseverance of that belief. From the deliberative stance, I may, as Moran puts it, "address myself to the question of my state of mind in a *deliberative* spirit, deciding and declaring myself on the matter, and not confront the question as a purely psychological one about the beliefs of someone who happens also to be me" (2001, p. 63). The

idea of “deciding and declaring” as distinctive of the deliberative stance is in contrast to the mere “reporting” or “informing” role of the theoretical stance. In saying that the deliberative stance is a stance from which one *decides* about what to believe, the idea here is that beliefs, as well as many other attitudes, are, at least often enough, subject to rational evaluation and can be formed or discarded in response to rational activity. For example, I can form a belief as the result of considering evidence.

Furthermore, I take my beliefs to be the kind of things that I, in general, have some obligation to defend. In saying that the deliberative stance is a stance from which one *declares* what it is one believes, Moran is noting the possibility for self-ascriptions of belief to be expressions of, rather than mere descriptions of, one’s commitments. Thus, when Moran claims that “there is indeed a dynamic or self-transforming aspect to a person’s reflections on his own state, and this is a function of the fact that the person himself plays a role in formulating how he thinks and feels,” Moran is noting that a person’s consideration of the reasonableness of her own beliefs can, and often enough does, result in a change in what those beliefs actually are (2001, p. 59). It is in these ways that the deliberative stance is an agential stance; it is a stance from which I consider my beliefs as the kinds of things that should be evaluated according to the evidence, where my evaluations can effect changes in what it is I actually do believe, and where my statements about what I believe can express, rather than merely report, my own commitments.

Why get into the weeds of how to characterize these different stances when our aim was to consider epistemic self-assessments? Because, with this distinction, we are now poised to ask which kind of reflection is the relevant kind required for the epistemic self-assessments that some have thought to be so important.

### 3.

In this section, I examine Ernest Sosa's account of the importance of reflective knowledge and fully apt knowledge. I offer four reasons for why I am choosing to focus on Sosa's account, rather than on internalism more generally. First, Sosa offers what is arguably one of the most sustained defenses of reflection available in contemporary epistemology. Second, Sosa takes his account to be able to accommodate many of the internalist intuitions about knowledge, such as those that occupied our attention in section 1 (2015; 1992). Third, because Sosa develops his account of the importance reflection *in addition to*, rather than *in opposition to*, an externalist account of knowledge, we can approach the topic from what is perhaps more neutral ground (neutral between internalism and externalism, anyway). Fourth, agency has played an increasingly important role in Sosa's most recent work, thus it will be instructive to compare my account with his.

I begin with a summary of Sosa's view. Initially, Sosa presented a two-tier account of knowledge: *animal* and *reflective*. Animal knowledge requires apt belief, which is true belief that results from the exercise of a competency or skill one possesses. Aptness is an externalist notion: one can have a true belief that is the result of an exercise of one's competency, and thereby have animal knowledge, without being aware that one's belief has this particular epistemic credential (i.e. that one's belief is apt). But according to Sosa, (mere) animal knowledge is insufficient to account for the kind of knowledge that *mature humans* possess, remarking that "Human knowledge is on a higher plane of sophistication, however, precisely because of its enhanced coherence and comprehensiveness *and its capacity to satisfy self-reflective curiosity*. Pure reliabilism is questionable as an adequate epistemology for such knowledge." (1992, p. 95, emphasis mine). It is important to note that the inadequacy of animal knowledge that Sosa identifies is not specific to animal knowledge *per se*, but arises insofar as animal knowledge is a

type of externalist (reliabilist) knowledge. I will discuss coherence in more detail later. It is enough at this point to note that the capacity to satisfy self-reflective curiosity requires being able to achieve some kind of *self-knowledge about one's own epistemic status* insofar as the object of inquiry is an assessment of the epistemic status of one's own beliefs.

One might think that the cure for whatever ails externalist knowledge is *reflection*. After all, it is here that Sosa introduces reflective knowledge as the kind of knowledge that has “enhanced coherence, comprehensiveness, and the capacity to satisfy self-reflective curiosity” (1992). However, Sosa's account of reflective knowledge is disappointing. For one thing, Sosa identifies reflection in a somewhat idiosyncratic way as a higher-order meta-belief. On his account, reflective knowledge is apt belief at the second-order, requiring not only apt beliefs, but also apt meta-beliefs that one's first-order beliefs are apt (2007, p. 108). But, as others have pointed out, having higher-order beliefs does not guarantee that a subject is aware of her beliefs at either the higher-order or lower-order level (Kornblith, 2009; Rödl, 2007). Thus, the conditions for reflective knowledge can be satisfied by a belief that is second-order but, nevertheless, is still externalist insofar as it is altogether unavailable to the subject's own point of view. And it's hard to see how higher-order externalist knowledge would be able to address internalist concerns. After all, reiterating externalist knowledge at a higher-order does not, in virtue of that fact alone, somehow transform the knowledge into internalist knowledge. If reflection is what's needed, reflection understood as meta-beliefs will not suffice.

What kind of reflection would do the trick? It sounds like the trouble here is that Sosa's account allows that reflection may be unavailable to consciousness. This is, I think, a common response to the line of objection I have been developing above. The problem with it, however, is that even conscious reflection, understood as awareness *of* the epistemic credentials of one's

belief, is not sufficient to provide the goods that Sosa claims are the result of reflective knowledge. Those goods require agent awareness, or so I shall argue.

Sosa identifies a variety of goods that are obtained when one has reflective knowledge, and I will not be able to address them all here. Here is a small sample of some of what reflective knowledge supposedly provides:

Human knowledge is on a higher plane of sophistication, however, precisely because of *its enhanced coherence and comprehensiveness* and its *capacity to satisfy self-reflective curiosity*. Pure reliabilism is questionable as an adequate epistemology for such knowledge. (1992, p. 95, emphasis mine)

“Why the pride of place for reflective knowledge?...One answer is to be found in the *special bearing of reflective knowledge on the understanding and coherence* dear to intellectuals...It is the ideal of reflective knowledge that best explains the traditional attraction and importance of skepticism...Finally, *reflection aids agency*, control of conduct by the whole person, not just by peripheral modules.” (2004, p. 291-292, emphasis mine)

Animal knowledge is essentially apt belief, as distinguished from the more demanding reflective knowledge... requiring not only apt belief but also *defensibly apt belief*. (2007, p. 24, emphasis mine).

Several epistemic values stand out:

- (a) Truth: we would rather our beliefs were true than not true, other things being equal.
- (b) Safety: we would prefer that not too easily would our beliefs be false.
- (c) Understanding/explanation: often we would like not only to know a given thing, but also to understand it, to have an explanation. (And this leads to the next item.)
- (d) Coherence: we would prefer that our minds not house a clutter of mere facts sitting there loose from one another.
- (e) Finally, we are often interested not only in *having* the truth but in *discovering* it, which involves not just being visited with the truth by sheer happenstance or through some external agency, but to arrive at the truth through our own intelligent doings, by relying on our own reliable abilities, skills, and faculties. (2009, p. 137)

Early on, Sosa identifies coherence and satisfying self-reflective curiosity as the central concerns; to these, he adds defensibility and aiding agency as being of special significance. But it is not especially clear how each of these things relate to one another. Sosa gives us a bit of a

grab-bag of epistemic goods, each of which is supposedly somehow made available through reflection. I will proceed by considering three of the goods in this grab-bag: coherence, defensibility, and aiding agency. The goal is to see whether these goods require agent-awareness of one's beliefs (here understood as equivalent to being able to take the deliberative stance), or if they can be satisfied when one has only awareness of one's beliefs and their epistemic credentials (one has only the theoretical stance). In this section I will examine *coherence* and *defensibility*, and I argue that each comes in a weaker and stronger version. I intend to show that reflection (in either sense) is not required for the weaker version, thus the weak versions cannot motivate reflective knowledge. The stronger version of coherence and defensibility are important goods, but each requires agent-awareness. As far as I can tell, satisfying self-reflective curiosity does not require the theoretical stance, and thus I will not discuss it in detail here. The argument of this entire paper is, in a certain sense, an argument that satisfying one's self-reflective curiosity (by being aware of the epistemic status of one's belief) plays a much smaller role in motivating internalism and Sosa's higher-orders of knowledge than is commonly recognized. If someone could show that agent concerns are also in effect here, it would only strengthen my case. There is also a third good I consider, that of *aiding epistemic agency*. However, since Sosa develops his account of the significance of agency in his account of *fully apt knowledge*, rather than reflective knowledge, I will examine agency in the following section where I discuss fully apt knowledge.

Coherence: Sosa claims that reflective knowledge has a "special bearing" on the coherence that is "dear to intellectuals" (2004). We might wonder how reflection contributes to the coherence of one's beliefs. One possibility is that a person is able to, through reflective

management of her beliefs, achieve a degree of coherence that she would be unlikely to achieve otherwise. For instance, suppose that there is a certain kind of incoherence in the beliefs of a subject S. Perhaps S believes that  $p$  and also that  $not-q$ . However, it is also the case that  $p$  entails  $q$ , although it is a matter S has never considered. It is plausible to think that S would benefit a great deal from careful reflection on  $p$  and  $q$  and the relation between them. After all, were she to reflect, she would realize that  $p$  entails  $q$ , and that she must either reject  $p$  or reject  $not-q$ . So perhaps Sosa's idea is that when someone has (rightly) judged that her beliefs are rational, those beliefs have a kind of guarantee of coherence that unscrutinized beliefs do not possess.

But the mere fact that a coherent set of beliefs is better than an incoherent set of beliefs doesn't yet explain why reflective belief is better than unreflective belief (Grimm, 2016 also makes this point). To get the conclusion that reflective belief somehow improves the coherence of one's beliefs, it must also be the case that unreflective beliefs are, on average, less coherent than reflectively scrutinized beliefs. However, conscious reflection is certainly not unique as a way to manage—and presumably increase—the coherence of one's beliefs. The coherence of our belief systems may be monitored through subconscious processes that generally operate quite well without reflective scrutiny.

Perhaps, then, the idea is that beliefs that survive reflective scrutiny have been “reviewed”, as it were, by an additional reliable process. If this is what Sosa has in mind, then Sosa's claim has an empirical component. Does the empirical evidence show that reflective management of one's beliefs reliably (not infallibly) increase the likelihood that one's beliefs are coherent? Hilary Kornblith has argued extensively that it does not (2012; 2013; 2016). The empirical evidence does strongly suggest that people often are not very good at accurately identifying the sources of their beliefs, or as Kornblith puts it, “Subjects are ignorant of the

actual source of their beliefs, and reflection is incapable of revealing it to them” (2016, p. 170). But Kornblith doesn’t seem to consider that the relevant reflective management of one’s beliefs may not be a descriptive project. Suppose I wish to find out if I have good reasons for believing that the crowd at last year’s inauguration ceremony was much smaller than the crowd at the inauguration ceremony four years before. I may be interested in evaluating the causal history of my belief, but not necessarily. It is, I think, more likely that I am interested in what sort of evidence there is for this belief in order to determine if it is rationally defensible. Good reflective management of one’s beliefs is focused on reasons and evidence for a belief, not merely developing an accurate causal history (for a related discussion, see Pust, 2014).

But additional reliability doesn’t seem to be all Sosa has in mind here. His claim is not just that one’s beliefs are *more* coherent when they result from reflection. He claims that being aware of the coherence of one’s beliefs results in a distinct *type* of justification:

For it is by adding interestingly to the coherence of one’s picture of the world and one’s place in it that one is able to gain a further measure of distinctive, epistemically valuable justification for one’s own empirical beliefs, a measure of justification that goes beyond the mere reliability of those beliefs, a reliability we can see to follow from how we must acquire contents and form beliefs. The additional measure of justification goes beyond any delivered by sheer reliability, and does so by bringing to consciousness a self-founded account of how our nature and emplacement yield such reliability. (2009, p. 131)

It seems, then, that Sosa does not think that the value of reflection is to be found simply in the fact that it is a reliable process that confers additional *reliabilist* justification. What other kind of justification is it? Sosa’s answer is that there is some (here unspecified) connection between a person’s conscious view and that person’s agency:

A belief constitutive of reflective knowledge is a higher epistemic accomplishment if it coheres properly with the believer’s understanding of why it is true (or apt) and of how the way in which it is sustained is reliably truth-conducive. Cohering thus within the believer’s perspective is, moreover, not

irrelevant to a belief's being deeply attributable to the believer's epistemic agency." (2009, p. 138)

But now, in order to take seriously the claim that reflection confers *additional* coherence in *virtue of being the result of* active management of one's beliefs, the type of reflection Sosa has in mind here cannot be merely an awareness *of what* one's beliefs and their epistemic credentials, for then there is no reason to think that this knowledge should result in any adjustment of her beliefs. If a person can identify incoherence among her beliefs but her beliefs are not responsive to her own assessments (no belief revision, for instance), then reflection cannot provide additional coherence. The idea that beliefs that pass reflective scrutiny have, for that very reason, a special epistemic status such that it is a belief that "would survive deep reflection by the subject in light of her deepest epistemic standards" (2009, p. 37) is plausible, to the extent that it is plausible at all, only against a background assumption that the person doing the reflecting is also able to make adjustments in what she believes as a result of her reflection. Otherwise, there would be nothing special about a belief surviving deep reflection; one's reflections simply wouldn't be the kind of thing that could result in the alteration of one's beliefs. If Sosa is right that reflection is crucial to epistemic agency, the relevant kind of reflection must involve agent-awareness.

Defensibility: A second reason that Sosa gives for the value of reflective knowledge is that, unlike animal knowledge, reflective knowledge is *defensible knowledge*:

Nevertheless, a mere thermometer reaction to one's environment cannot constitute the best human knowledge, regardless of whether that reaction is causally mediated by experience. It is not enough that one respond to seeing white and round objects in good light with a 'belief' or 'proto-belief' that there is something white and round. Suppose one asks oneself 'Do I know that this is right and round' or 'Am I justified in taking this to be white and round?' and one has to answer 'Definitely not' or even 'Who knows? Maybe I do know, maybe I don't; maybe I'm justified, maybe I'm not.' In that case one

automatically falls short, one had attained only some lesser epistemic status, and not any ‘real, or enlightened, or reflective’ knowledge. The latter requires some awareness of the status of one’s belief, some ability to answer that one does know or that one is epistemically justified, and some ability to defend this through the reliability of one’s relevant faculties when used in the relevant circumstances. (Sosa 2009: 153)

In fact, Sosa has used the notion of defensibility to characterize reflective knowledge, by referring to reflective knowledge as “defensibly apt belief” (2007, p. 24). Here we might ask again, as we did with coherence, what does reflection have to do with defensibility? Answering this question requires saying a bit more about what it is for a belief to be defensible. Habib and Lehrer offer this characterization of Sosa’s requirement of defensibility: “let us call this constraint, one requiring that a target belief that coheres with a system of beliefs can be defended by the system against objections to the belief, the *defensibility constraint*” (2004, p. 109). Notice that in their description of defensibility, Habib and Lehrer emphasize the relation between a belief and a *system* of supporting beliefs. There is certainly something plausible about the idea that a belief is defensible when other beliefs within the system can be shown to support that belief. And if I am asked why I believe that *p* is true, my ability to defend my belief that *p* does seem to require my ability to identify other things I believe that count in favor of *p*.

But there is also something odd about this phrasing. Note that in their description, the defender is *the system*. It sounds, at this point, like defensibility amounts to a kind of coherence that includes relevant evidential support relations. However, having a coherent system in which a belief is appropriately supported by other beliefs in the system isn’t sufficient for defensibility, for it is possible to *show how a belief coheres within a supporting system of beliefs* without *defending one’s belief*. For example, a professor might give a writing assignment instructing her students to defend a particular ethical position, and some students may do so adequately, or even masterfully, while at the same time disagreeing with the view they defend. (Presumably this

happens frequently in activities such as formal debates, playing certain games, and even where defending a certain position is required for doing one's job responsibly.)<sup>ii</sup> Furthermore, the system must be aware of these relations. Apt meta-beliefs of which one is unaware would not provide defensibility, for the subject is in no position to defend her beliefs if she is unaware of anything at all that supports them.<sup>iii</sup> Is it, then, the case that being aware *that I believe p* and being aware *that p has certain epistemic credentials* (say, aptness) makes my belief that *p* defensible?

Notice that the preceding paragraph began with a discussion of *systems*, and concluded with a discussion of *subjects*. This shift is of crucial importance. Consider that a defense of *my* beliefs is a defense of *my* commitments, of the things I take to be true (or that I take to be likely enough to be true). As such, defending my beliefs requires recognizing that my commitments are in question, that I can be at least in some sense held responsible for the things I believe, and that beliefs are the kinds of things I might be called upon to defend. Of course, often in giving a defense we do identify what supports what. But while I may report on the evidential relations among the various beliefs in my belief system, a defense of my beliefs requires more than a mere report. This is the case even if it is a true report of the things I do believe, including the evidential relations among the relevant beliefs in the system. Otherwise, defending one's belief would be merely an exercise in identifying what follows from what. To put it one way, *systems* don't defend beliefs, even if systems of beliefs are what we appeal to when making a defense. And *defending my belief* requires more than being able to defend a belief that also, as a matter of fact, happens to be my own. Defending one's own belief implicates one's commitments, and it is precisely this notion of commitment that is not available from the theoretical stance, where one

may indeed accurately report on the evidential relations among one's beliefs. That is to say, defending one's own beliefs requires agent-awareness of those beliefs.

#### 4.

Thus far, I have argued that the sort of coherence and defensibility Sosa seems concerned about in his arguments for the importance of reflective knowledge both require agent-awareness of one's belief. Without agent-awareness, we cannot make sense of the notion of there being additional justification conferred by a belief's surviving reflective scrutiny. So, too, without agent-awareness, the notion of defending one's belief amounts to a psychological self-report, rather than an activity where one's commitments are crucially relevant. In this final section, I turn to a consideration of the idea that reflection is valuable because of the role it plays in agency (including practical agency, but I shall focus on the epistemic). Again, Sosa argues that reflection is important because it somehow "aids agency", and that we desire to discover the truth, to arrive at it "through our own intelligent doings, by relying on our own reliable abilities, skills, and faculties" (2009, p. 137). These remarks about the importance of agency merely gestured toward the view that Sosa develops in detail in his most recent work on *fully apt knowledge*.

Sosa now claims that "The importance of the reflective is not explained fully until we see what really matters: namely, that the aptness on the first order be attained under the guidance of the second-order awareness" (2015, p. 85). Thus, Sosa introduces a third category of knowledge: knowledge is *fully apt* when one has a higher-order belief that one's first-order belief is apt, such that one's first-order belief both *manifests* and is *guided by* the person's apt reflective activity (Sosa, 2011; 2015). It is *fully apt knowledge* that now stands at the apex of the hierarchical epistemic triad, with reflective knowledge and animal knowledge occupying the lower rungs (respectively). The value of reflective knowledge, over and above mere externalist knowledge, is

actually derived from the fact that reflection is required for fully-apt knowledge. We turn our focus, then, to considering why fully apt knowledge is purportedly superior to both animal knowledge and reflective knowledge.

Here Sosa's answer to this question has everything to do with the fact that fully apt knowledge is agential. Sosa provides two arguments. The first appeals to the general structure of performances. According to Sosa, when one considers our evaluation of performances, taken as a type, fully apt performances occupy the highest rung on the evaluative scale. That is to say, of all performances, only fully apt performances are those that are not deficient, or as Sosa puts it, "any performance suffers if it is not fully apt" (2015, p. 87). Sosa defends this claim through detailed analysis of various particular cases of performances. I will refer those who wish to consider the cases in detail to Chapter 1 of *Judgment and Agency* (2015), but to sum, the key idea is that fully apt performances are of highest value precisely because it is only then that the performance (or its result) is fully creditable to the agent. Thus the value of fully apt knowledge can be explained by reference to the fact that fully apt knowledge is the result of fully apt epistemic performances.

The second argument focuses on the specific value of fully apt knowledge. It is here that we might attempt to locate the goods of enhanced coherence and defensibility that reflective knowledge purportedly affords. Fully apt beliefs are guided by one's higher-order evaluations of one's beliefs, and this guidance almost certainly will require attending to coherence and to the epistemic credentials of one's beliefs. Also, Sosa points to the importance of fully apt knowledge (which he sometimes refers to as "judgmental" knowledge) in social coordination:

Genuine judgment is affirmation in the endeavor to affirm with apt correctness...Such conscious judgments and the corresponding beliefs are important to a social species for which information sharing and collective deliberation are essential, as is cooperation more generally. This largely accounts

for the great value of the judgmental knowledge constituted by such acts or dispositions. Affirmation and judgment (as defined) are crucial to the various forms of human cooperation. (2015, p. 66)

By considering its social importance, Sosa extends the argument beyond individualistic concerns, such as answering first-personal questions about the epistemic status of one's beliefs (satisfying one's self-reflective curiosity). His attention on the social importance of knowledge is, I think, quite important. Actually, both ways of defending fully apt knowledge seem plausible. However, note that both of these defenses locate the value of fully apt knowledge in agential activity. It is agent-credit that gives fully apt performances their value, and social activities Sosa emphasizes require commitment to belief that is available only from the deliberative stance.

Although my aim in this paper is to illuminate motivations for internalism more generally, I will briefly address how the distinction between mere reflection and agent-awareness relates to Sosa's current view. As far as I can tell, the distinction between mere reflection and agent-awareness can fit nicely into Sosa's account. In fact, agent-awareness is required for the intentional pursuit of the aim of apt belief and for beliefs to be guided by one's estimation of whether or not one's beliefs are apt. Sosa identifies two ways in which agency is implicated. First, fully apt knowledge requires engaging in an intentional action: namely, the (conscious) pursuit of apt belief. This requirement does not entail the explicit verbal expression of the intention (one might hike up a mountain intentionally without having the explicit thought, "I intend to make it to the top".) But the proper description of one's action must be to determine whether one's belief is apt, or as Sosa puts it, "Affirmation in the endeavor to answer a whether-question correctly and also competently, reliably enough, even aptly" (2015, p. 55). Second, agency is exercised at the higher-order level by the endorsement or withholding of endorsement from one's seemings. The talk of seemings is contentious, but as Sosa draws the distinction,

‘seemings’ are roughly *how things appear to us*. These seemings are not always coercive; at least in many circumstances, there is the possibility of withholding higher-level assent of those seemings as factive. Crucially, one must determine how much assent is warranted, given the evidence one has that one’s seemings are, in fact, apt. Optical illusions illustrate the point. It is possible, for instance, for it to seem that two lines are not equal, while at the same time be aware that this seeming is an illusion. When one is aware that one is looking at an illusion, one may withhold assent from one’s seeming, even while the seeming persists. Both of these sources of epistemic agency require agent-awareness. It is from the deliberative stance, not the theoretical, that one sees one’s beliefs as able to be guided by one’s own reflective activity *at all*, much less that such reflective activity would be normatively required (by the demands of rationality or something of that sort). Also, to endorse a seeming is to implicate one’s commitments; one cannot endorse and remain in the position of an impartial observer.

Furthermore, I agree with Sosa that reflection has the significance it does because reflection enables agency. But the details work out a bit differently in my story of how reflection enables agency. How significant are those differences? Here I offer two thoughts. First, on my view it wasn’t a *mistake* to identify reflection, or a higher-order perspective as sufficient for a particular kind of epistemic agency. An objection to Sosa raised by Stephen Grimm may help to bring out this difference. Grimm argues that “Sosa misidentifies the point at which autonomy enters the story. It is not by reflecting on the reliability of our faculties that we acquire intellectual autonomy, or that our beliefs become properly creditable to us. Rather, it is by taking our ‘given’ ends as believers to be good or worthwhile that we acquire this autonomy.” (2016, pp. 191-2) How might Sosa respond? First, whether or not Sosa once held that intellectual autonomy was held in virtue of reflecting on the reliability of one’s faculties, this is not his

current view. In fact, Grimm's suggestion actually sounds a good deal like Sosa's claim that fully apt belief is agential in virtue of it resulting from the pursuit of apt belief as an intentional aim. But on my view, Sosa's original claim *that intellectual autonomy is acquired by (or I prefer, in) reflecting* was wrong only if one is referring to mere reflection. Intellectual autonomy is present whenever one has agent-awareness. Second, when considering one's beliefs from the deliberative stance, one is already located as an agent, even if one has not yet adopted any particular aim. On my view, epistemic agency does not depend on adopting any particular aim.

#### 4.

I began with the claim that a key internalist motivation is to obtain a particular kind of self-knowledge, knowledge of the epistemic credentials of one's belief. I then argued that those who think there is something missing from mere externalist knowledge will almost certainly be similarly dissatisfied with mere reflection (from the theoretical stance). Again, a quotation from Sosa illustrates the point. Sosa once remarked that unreflective knowledge is unsatisfactory because leaves one groping in the dark: "The Pyrrhonists reject ... externalism because it dignifies mere 'groping in the dark' with the title of knowledge. The Pyrrhonists highlight enlightened knowledge, acquired and sustained in awareness of one's epistemic doings. Only this is 'knowledge' worthy of the title" (1997, p. 242). Other virtue epistemologists have made a similar point. For instance, Linda Zagzebski says "The value of the truth obtained by a reliable process in the absence of any conscious awareness of a connection between the behavior of the agent and the truth he thereby acquires is no better than the value of the lucky guess" (1996, p. 304). However, if all we can do is assess our epistemic standing without having those

assessments make a difference to what it is we believe, then even conscious reflective knowledge turns out to be only so much *groping in the light*.

It is, I submit, largely for this reason that Sosa now thinks that one's lower-order apt belief must be guided by one's higher-order reflection. But much of the discussion of the importance of reflection in epistemology still entangles concerns about an awareness *of* the epistemic status of one's beliefs—something that does not require the ability to take any sort of agential stance toward one's own attitudes—with concerns that do require the ability to take an agential stance toward one's beliefs. There is, of course, a great deal more work to be done here, but if I'm right, sorting out mere reflection and agent-awareness is one promising way to start.

## Bibliography

- Battaly, H. (2008). "Virtue Epistemology". *Philosophy Compass*, pp. 639–663.
- Bergmann, M. (1997). Internalism, Externalism and the No-Defeater Condition. *Synthese*, 110, 399-417.
- Bergmann. (2006). *Justification Without Awareness: A Defense of Epistemic Externalism*. (Oxford: Clarendon Press)
- Bonjour, L. (1980). Externalist Theories of Empirical Knowledge. *Midwest Studies in Philosophy*, 5, 53-73.
- Bonjour, L. (2001). The Indispensability of Internalism. *Philosophical Topics*, 29, 47-65.
- Coppenger, B. (2016). Traditional Internalism: An Introduction. (In B. Coppenger & M. Bergmann (Eds.), *Intellectual Assurance: Essays on Traditional Epistemic Internalism* (pp. 1-20). Oxford: Oxford University Press.)
- Fumerton, R. (2004). Achieving Epistemic Ascent. (In J. Greco (Ed.), *Ernest Sosa and His Critics* (pp. 72-85). Oxford: Blackwell Publishing Ltd.)
- Greco, J. (2003). Knowledge as Credit for True Belief. (In M. DePaul & L. Zagzebski (Eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (pp. 111-134). Oxford: Oxford University Press.
- Grimm, S. R. (2001). Ernest Sosa, Knowledge, and Understanding. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 106, 171-191.
- Grimm, S. R. (2016). The Value of Reflection. (In M. Á. F. Vargas (Ed.), *Performance Epistemology: Foundations and Applications* (pp. 183-195). New York: Oxford University Press.)
- Habib, A., & Lehrer, K. (2004). Sosa on Circularity and Coherence. (In J. Greco (Ed.), *Ernest Sosa and His Critics* (pp. 106-111). Oxford: Blackwell Publishing Ltd.)
- Kornblith, H. (2009). Sosa in Perspective. *Philosophical Studies*, 144, 127-136.
- Kornblith, H. (2012). *On Reflection*. (Oxford: Oxford University Press)
- Kornblith, H. (2013). Naturalism and the First-Person. *Proceedings of the American Philosophical Association, Romanell Lecture*, 107-126.
- Kornblith, H. (2016). Epistemic Agency. (In M. Á. F. Vargas, *Performance Epistemology: Foundations and Applications* (pp. 167-82). New York: Oxford University Press)

- Lackey, J. (2007). "Why We Don't Deserve Credit for Everything We Know". *Synthese*, 158, 345-361.
- Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. (Oxford: Oxford University Press)
- Lehrer, K. (1990). *Theory of Knowledge*. (Boulder: Westview Press)
- Moran, R. (1999). The Authority of Self-Consciousness. *Philosophical Topics*, 26, 179-200.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. (Princeton: Princeton University Press)
- Perry, J. (1979). The Problem of the Essential Indexical. *Nous*, 13, 3-21.
- Pust, J. (2014). Critical Notice of Hilary Kornblith's *On Reflection*. *Episteme*, 11, 53-61.
- Rödl, S. (2007). *Self-Consciousness*. (Cambridge: Harvard University Press)
- Sosa, E. (1992). *Knowledge in Perspective*. (Cambridge: Cambridge University Press)
- Sosa, E. (1997). How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes. *Philosophical Studies*, 85, 229-249.
- Sosa, E. (2004). Human Knowledge, Animal and Reflective: Reply to Robert Audi, John Greco, and Hilary Kornblith. (In J. Greco (Ed), *Ernest Sosa and His Critics* (pp. 290-292). Oxford: Blackwell Publishing Ltd.)
- Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. (Oxford: Oxford University Press)
- Sosa, E. (2009a). Knowing Full Well: The normativity of beliefs as performances. *Philosophical Studies*, 142, 5–15.
- Sosa, E. (2009b). *Reflective Knowledge: Apt Belief and Reflective Knowledge, Volume II*. (New York: Oxford University Press)
- Sosa, E. (2011). *Knowing Full Well*. (Princeton: Princeton University Press)
- Sosa, E. (2015). *Judgment and Agency*. (New York: Oxford University Press)
- Turri, J. (2013). Bi-Level Virtue Epistemology. (In J. Turri (Ed.), *Virtuous Thoughts: The Philosophy of Ernest Sosa* (pp. 147-164). New York: Springer.)
- Zagzebski, L. (1996). *Virtues of the Mind*. (Cambridge: Cambridge University Press)

---

<sup>i</sup> On Moran's view, occupying the theoretical stance involves a kind of estrangement; the best examples of occupying merely the theoretical stance are those in which something has gone awry. Moran uses the distinction to address cases of weakness of will. We might suppose that someone who tends to overeat is asked whether or not he believes he will remain on his diet over the holidays. This person might say that he intends to remain on his diet, while at the same time, given his history of overeating, he also doubts that he will remain on his diet. This doubt that he will remain on his diet is an assessment that he makes as a result of considering his history and observing that he has never yet remained on a diet over the holidays. When his assessment of what he believes he will do depends on evaluating his relevant behavior, rather than arising from his own intentions, he is occupying the theoretical stance toward his belief that he will, despite his present intentions, not remain on his diet.

<sup>ii</sup> For example, consider a case raised by Jennifer Lackey (2008) in which a teacher who believes the universe was created by intelligent design might nevertheless recognize that it is his responsibility to teach evolution.

<sup>iii</sup> Here one might also press Sosa on the strength of this requirement, especially in regard to skeptical challenges. For instance, does defensibility require being able to defend one's belief against all challenges, or only some? If the former, then is any belief defensible? If the latter, then which challenges are the relevant ones, and why does reflective knowledge require defensibility only to those? These are important questions, but getting clear on the scope of the requirement of defensibility is not our present concern.