

## ***Sosa's bi-level virtue epistemology***\*

JOHN TURRI

[john.turri@gmail.com](mailto:john.turri@gmail.com)

Ernest Sosa has long defended *bi-level virtue epistemology* on the grounds that it offers the best overall treatment of epistemology's central issues. A surprising number of problems "yield to" the approach (Sosa 1991: 9). Sosa applies bi-level virtue epistemology to diagnose and bypass ongoing disputes in contemporary epistemology, including the disputes between foundationalists and coherentists, and between internalists and externalists. He also invokes it to explain the nature of epistemic value and the assessment of intellectual performance, to define knowledge, and to defend against skeptical challenges, among other things. Although the two aspects of Sosa's view, virtue theory and bi-level epistemology, are intimately connected, they are nonetheless conceptually distinct and make isolable contributions to Sosa's overall project. This chapter will focus primarily on contributions made by virtue theory, and secondarily on contributions made by bi-level epistemology, where they are especially relevant to appreciating the limits of the work done by virtue theory in Sosa's epistemology.

---

\* This is a draft of work in progress. Comments welcome. Please don't cite, quote or refute without permission. This research was supported by the Social Sciences and Humanities Research Council of Canada.

### 1. *Foundationalism and coherentism*

The great Scottish philosopher David Hume once argued that *ambiguity* is the best explanation for persistent disagreement between parties to a longstanding debate. Wrote Hume,

From this circumstance alone, that a controversy has been long kept on foot, and remains still undecided, we may presume that there is some ambiguity in the expression, and that the disputants affix different ideas to the terms employed in the controversy. (Hume 1748: section 8.1)

But beginning with his work in the late 1970s, Sosa takes a different approach to the debate between foundationalists and coherentists over the structure of knowledge. (Indeed, Sosa takes this different approach to a number of longstanding disputes in philosophy.) Rather than assuming the sides are talking past one another, Sosa suggests that each side has identified part of the truth, but missed out on the bigger picture.

In an area so long and intensively explored it is not unlikely that each of the main competing alternatives has grasped some aspect of a many-sided truth not wholly accessible through any one-sided approach. The counsel to open minds and broaden sympathies seems particularly apt with regard to basic issues so long subject to wide disagreement. (Sosa 1991: 78)

Sosa proposes that virtue epistemology can capture what is attractive in both foundationalism and coherentism. He makes this case most completely in his famous paper “The Raft and the Pyramid”

(Sosa 1991: ch. 10), so I will focus on it.<sup>1</sup>

A key idea in Sosa's discussion is *supervenience*, and in particular the supervenience of the evaluative on the nonevaluative. It is widely accepted that all evaluative properties supervene on nonevaluative properties. To understand why this view seems so plausible, let's first clarify what we mean by 'supervene', 'evaluative' and 'nonevaluative'.

Supervenience can be neatly defined. Supervenience is a relation between two classes of properties. Let 'A-properties' and 'B-properties' name two distinct sets of properties. The A-properties supervene on the B-properties just in case no two things can differ in their A-properties without also differing in some of their B-properties. Put otherwise, there can't be an A-difference without a B-difference. When the A-properties supervene on the B-properties, we call the A-properties *supervenient* and the B-properties *subvenient* or *base properties*. It is also implied that the A-properties obtain because of or in virtue of the B-properties.

It isn't easy to informatively and uncontroversially define what counts as an evaluative property, but the following should suffice for present purposes. Evaluative properties are ones that feature centrally in evaluation, as when we judge something to be *right*, *wrong*, *proper*, *improper*, *good*, *bad*, *worthy*, *unworthy*, or the like. Nonevaluative properties are the ones that feature in what we might call a "neutral" description of something. For instance if I

---

<sup>1</sup> See also "The Foundations of Foundationalism" (reprinted in Sosa 1991: ch. 9) and "Epistemology Today: A Perspective in Retrospect" (reprinted in Sosa 1991: ch. 5).

hold forth a spade and say, “this is a spade,” then I have described it neutrally. I haven’t evaluated it or, as they say, “passed judgment” on it, although I have clearly classified it by placing it in the category of spades. By contrast if I say, “this is a good spade,” then I have gone beyond merely classifying it to evaluating it. I have described it, but not neutrally.<sup>2</sup>

Now we can see why it is widely assumed that the evaluative supervenes on the nonevaluative. First, if a spade is a good spade, then it isn’t just a brute fact that it’s good. There must be an explanation of why it’s good. And the explanation certainly seems to be that it’s good because of its durability, strength, balance, comfortable grip, and other nonevaluative properties. Of course in some cases one evaluative property could explain another. For example it might be worthy to purchase because it’s good, but then its worthiness (to purchase) would still ultimately supervene on the nonevaluative properties that explain its goodness. Second, it also seems that two things identical in their nonevaluative properties must also be identical in their evaluative ones. Consider how absurd it would be to maintain that although two spades were *indistinguishable* in terms of their strength, durability, balance, and so on, one of them is nevertheless good while the other isn’t. Surely such an outcome is impossible.

So all evaluative properties supervene on nonevaluative proper-

---

<sup>2</sup> I don’t intend to equate describing something *neutrally*, as I use that term here, with describing it *objectively* or *factually*. For all I’ve said, reality might not be neutral, and evaluative descriptions might denote objective facts. For more on Sosa’s view of objectivity in matters of value, see chapter [[]] of this volume.

ties. And epistemic properties, including justification and knowledge, are evaluative properties. So epistemic properties, including justification and knowledge, supervene on nonevaluative properties. Call this *the epistemic supervenience thesis*.<sup>3</sup>

Sosa calls epistemic supervenience the “lowest” or most basic grade of “formal foundationalism” about epistemic properties. All that supervenience requires is a nonevaluative basis which guarantees that the belief is knowledge. This leaves open what that nonevaluative basis is. A higher grade of formal foundationalism accepts the epistemic supervenience thesis, and further maintains that the subvenient base properties “can be specified in general.” The highest grade of formal foundationalism accepts the epistemic supervenience thesis, and further maintains that the subvenient base properties can be simply and comprehensively specified.

Interestingly, coherentism and foundationalism, as standardly defined, are both forms of formal foundationalism. They disagree merely about what the base properties are. Coherentists say the base property is *coherence among a set of beliefs*. By contrast, foundationalists say it is *being grounded in perception* (the empiricist branch of foundationalism) or *being grounded in rational insight* (the rationalist branch), along with some appropriate mix of *introspection* and *memory*.

Sosa argues that this way of looking at epistemic properties

---

<sup>3</sup> For detail on variations of the epistemic supervenience thesis, see the entry on epistemic supervenience in *A Companion to Epistemology*, 2<sup>nd</sup> edition, ed. Jonathan Dancy, Ernest Sosa, and Matthias Steup (Wiley-Blackwell, 2010).

sheds new light on the debate between coherentists and foundationalists, and ultimately suggests a way beyond it entirely. Start with coherentism. Some antifoundationalist arguments used by coherentists start to look suspicious. For example Laurence BonJour and Wilfrid Sellars both argue that a true belief's being reliably produced isn't enough to ground knowledge. The subject would also have to know that it was reliably produced, they argue, and this is part of what makes the belief count as knowledge.<sup>4</sup> But this is not a good criticism of foundationalism, Sosa thinks, because it conflicts with the epistemic supervenience thesis. The subvenient base properties must be nonevaluative, but *knowledge* is an evaluative property, so demanding knowledge in the subvenient base is illegitimate.

Similarly sometimes antifoundationalists argue that a belief doesn't count as knowledge unless you also know that you wouldn't easily be misled about the claim in question. But then your belief isn't foundationally justified after all, because it's partly grounded in other knowledge. But this isn't a good criticism because it too conflicts with the epistemic supervenience thesis. Again, demanding knowledge in the subvenient base is illegitimate.

Sosa also criticizes coherentism for reasons independent of supervenience. One problem especially stands out, namely, its inability to account for justified beliefs only minimally integrated into our overall set of beliefs. Imagine that you have a splitting headache. You believe that you have a headache, and you have several other

---

<sup>4</sup> [BonJour and Sellars references]

beliefs that cohere with this, such as the belief that you're in pain, that someone is in pain, and that you're presently aware of a headache. This is a nice coherent set of beliefs, and it's very plausible that you're justified in accepting all of them. So far, so good. But now Sosa asks us to imagine the following modified case, in which everything about you, "*including* the splitting headache," remains the same, except that we replace the belief that you have a headache with the belief that you *don't* have a headache, replace the belief that someone is in pain with the belief that someone *isn't* in pain, and replace the belief that you're aware of a headache with the belief that you *aren't* aware of a headache. Your beliefs in the modified case are just as coherent as they were in the original case, so coherentism entails that this set of beliefs is equally justified as the set in the original case. But it seems obvious that this set of beliefs isn't justified.

Even though coherentism's prospects look bleak, Sosa doesn't conclude that foundationalism wins. Contemporary foundationalists typically claim that true beliefs based on perception, introspection, memory and rational insight count as knowledge. So they typically include these sources when specifying knowledge's subvenient base properties. The problem is that this list lacks unity. It seems like a *mere* list of conditions. Why just those sources? Call this *the scatter problem* for foundationalism. The question becomes more pressing when Sosa asks us to imagine "extraterrestrial beings" whose basic belief forming processes are nothing like ours, but nevertheless work well in their native extraterrestrial environments.

The foundationalist might well have to add more principles to his list, making it look even more scattershot. It would be better, Sosa proposes, “to formulate more abstract principles that can cover both human and extraterrestrial foundations.”

This brings us to Sosa's positive proposal, the initial statement of his virtue epistemology. He draws inspiration from the revival of virtue theory in the field of normative ethics. According to this view, moral virtues are the primary source of ethical justification. An action is right because it is produced by morally virtuous dispositions, or excellences of moral character, such as honesty and courage. A morally virtuous disposition is a character trait that enables the agent to promote good outcomes, or at least outcomes good enough under the circumstances and compared to the available alternatives. Sosa draws an important lesson from this “stratification of justification.”

The important move for our purpose is the stratification of justification. Primary justification attaches to virtues and other . . . stable dispositions to act, through their greater contribution of value when compared with alternatives. Secondary justification attaches to particular acts in virtue of their source in virtues or other such justified dispositions.

Sosa proposes that we adopt the same strategy for epistemic properties. Primary justification attaches to *intellectual* or *epistemic* virtues, “through their greater contribution toward getting us to the truth.” These virtues are dispositions to reliably believe the truth and avoid believing falsehoods. Secondary justification attaches to

individual beliefs for having been produced by the virtues. (Sosa often alternates between talk of “virtues” and “competences,” and between “dispositions,” “capacities,” “powers,” “faculties” and “abilities.” In almost every case, these are mere verbal variations and shouldn’t be taken to indicate a shift in the underlying view.)

Virtue theory helps us to understand what is right in both foundationalism and coherentism while avoiding their drawbacks. First consider coherentism. It is intellectually virtuous to accept a claim based on its coherence with other things we believe, because doing so reliably enough helps lead us to the truth. So believing based on coherence can enhance justification. But virtue epistemology doesn’t commit us to the view that coherence is the *only* thing required to gain justification or knowledge. Next consider foundationalism. We saw that it faces the scatter problem, a problem poignantly illustrated by the possibility of extraterrestrials who reliably form beliefs in ways utterly alien to us. Virtue epistemology offers a simple and principled explanation of why both our beliefs and the extraterrestrials’ beliefs are justified: they spring from intellectual dispositions that are, relative to their normal environments, reliable. Similarly we can explain why beliefs formed through perception, introspection, memory and rational insight all tend to be justified for us, despite their superficial disunity: our dispositions to trust these sources are virtuous.

It is crucial to Sosa’s view that the intellectual virtues have a nonevaluative basis, primarily in terms of how well they promote the acquisition of true rather than false beliefs. This is crucial be-

cause without it virtue epistemology can't respect the epistemic supervenience thesis. And if it violates the epistemic supervenience thesis, then much of Sosa's early motivation for it, at least, won't withstand scrutiny. An important question to consider, then, is whether the virtues do have a fully nonevaluative basis, or whether they instead have an irreducibly evaluative element.

Beginning in the early 1990s, another theme in Sosa's writings on foundationalism is that foundationalism needs virtue theory in order to account for foundational justification, or lack thereof, in even the simplest cases.<sup>5</sup> Sosa's favorite type of example for making this point involves a comparison between two different visual experiences: on the one hand, an experience of a well-lit, white *triangle* against a black background, and on the other hand, an experience of a well-lit white *dodecahedron* against a black background (Sosa 1991: 7ff; see also Sosa 2003a: ch. 7). For a normal human, the experience featuring a triangle justifies him in believing non-inferentially that he is currently experiencing a triangle, but the experience featuring a dodecahedron does not justify him in believing non-inferentially that he is currently experiencing a dodecahedron. 'Non-inferentially' here can be taken to mean roughly: at a glance, as opposed to counting the number of sides and inferring on that basis which type of polygon it is.

Why the difference between the two cases? The answer cannot simply appeal to how well the content of the experience matches the

---

<sup>5</sup> Precursors of this line of thought can be found earlier in Sosa's writings. E.g. see Sosa 1988: 171 (reprinted in Sosa 1991: cf. 127–8).

content of the relevant belief. After all, an experience featuring a dodecahedron matches the belief “this is a dodecahedron” just as well as an experience featuring a triangle matches the belief “this is a triangle.” Sosa explains the difference as follows. In the case of experiencing a triangle, normal humans have a “noninferential faculty that enables the formation of beliefs on the matter in question with a high success ratio” (1991: 9). In other words, they have an intellectual virtue that in normal circumstances makes them reliable at detecting at a glance whether they’re experiencing a triangle. This is why the experience justifies them in believing “this is a triangle.” By contrast, in the case of experiencing a dodecahedron, normal humans do not have a relevant reliable noninferential faculty or virtue. This is why the experience does not justify them in believing “this is a dodecahedron.” By contrast, if an especially gifted human had an ability to reliably detect, at a glance, that she was looking at a dodecahedron, then an experience of the experience of a dodecahedron would justify her in believing “this is a dodecahedron.”<sup>6</sup>

## 2. *Internalism and externalism*

Beginning with his work in the 1980s, Sosa applied virtue theory to

---

<sup>6</sup> Sosa’s solution to this problem for a time also relied on the claim that the belief in question was not only *virtuously* based on the relevant experience, but also *safely* (Sosa 2003a: 138–9); see Michael Pace’s discussion of the problem of the speckled hen in chapter [ ] of this volume. More recently, Sosa has abandoned any substantive safety requirement; see Sosa 2007 (especially chapters 2 and 5), my discussion below in section 3, and Juan Comesaña’s discussion of Sosa’s views on safety in chapter [ ] of this volume.

develop a theory of epistemic justification that accommodated the core intuitions of internalist epistemology within a broadly externalist framework.

More than one debate goes by the label “internalism versus externalism” in contemporary epistemology. All share one thing in common: they concern the nature and grounds of evaluative epistemic properties. The main such debate concerns epistemic justification. But even after we have narrowed the terrain to epistemic justification, there remain distinct senses in which one could be an “internalist.” For each sense of “internalism,” denying internalism in that sense makes you an “externalist” in that sense. Internalists claim that justification must be determined entirely by factors that are relevantly “internal,” and externalists deny this.

*Ontological internalism* says that all factors that help determine a belief's justification must be part of the believer's psychology.<sup>7</sup> *Ontological externalism* says that it's possible for justification to be at least partly determined by factors that are not part of the believer's psychology. *Access internalism* says that all factors that help determine a belief's justification must be unproblematically accessible to the believer. A typical access internalist understands “unproblematically accessible” to mean “available to the believer from the armchair, via introspection and a priori insight.”<sup>8</sup> *Access*

---

<sup>7</sup> I follow Sosa in calling it “ontological internalism” (Sosa 2003a: 146). (Compare Sosa 1991: 136: “What is internal in the right sense must remain restricted to . . . that which pertains to the subject's psychology.”) The view is also called “mentalism” in the literature, following Conee and Feldman 2001.

<sup>8</sup> Sosa also calls this “Chisholmian internalism”: “the view that we have spe-

*externalism* says that it's possible for justification to be at least partly determined by factors that are not unproblematically accessible to the believer.

Sosa aims to transcend the internal/external divide. A fully adequate epistemology must accommodate the intuitions motivating internalism, without going so far as to accept the internalist theses. The guiding thought, then, is that “externalism must find some way of doing justice to the appeal of epistemically internalist intuitions” (Sosa 2009: 44). In the remainder of this section, I'll first explain Sosa's treatment of ontological internalism, then I'll explain his treatment of access internalism. As we will see, Sosa thinks that although virtue theory can accommodate the intuitive basis of ontological internalism, bi-level epistemology is required to accommodate the motivation for access internalism.

The new evil demon thought experiment provides the most potent intuitive motivation for ontological internalism.

Compare yourself with a counterpart victim of the evil demon. Suppose the two of you indistinguishable in every current mental respect whatsoever; if you have a certain belief, so does your counterpart; if you would defend your belief by appeal to certain reasons, so would your counterpart; and vice versa. The two of you are thus point by point replicas in every current mental respect: not only in respect of mental episodes, but also in respect of deeply lodged dispositions to adduce reasons, etc. Must you then be equally epistemically justified, in some relevant sense, in each

---

cial access to the epistemic status of our beliefs . . . by means of armchair reflection” (Sosa 2003a: 145).

such belief that by hypothesis you share? . . . What could a difference in justification drive from? Each of you would have the same fund of sensory experiences and background beliefs to draw upon, and each of you would appeal to the same components of such cognitive structure if ever you were challenged to defend your belief. So how could there possibly be any difference in epistemic justification? (Sosa 2003a: 150)

Sosa agrees that it is “very implausible” that we are internally better justified than our twins are; we and our twins seem to be equally “internally justified” (1991: 132, 144). Sosa goes so far as to say that our twins are “internally justified in every relevant respect” (1991: 143), and that they might even be “flawlessly, and indeed brilliantly” internally justified in some respect (1991: 289). All this despite the fact that they are systematically deceived.

The challenge is to fully understand the internal justification that we and our twins share, but we can't do this by clinging to ontological internalism, Sosa argues. Ontological internalism inevitably misses dimensions of “internal epistemic excellence” and so “falls short” in explaining the full extent to which our twins are internally justified (Sosa 2003a: 148–9; compare Sosa 1991: ch. 8).

Consider several potential bases for supporting ontological internalism. First, ontological internalism might be supported on the grounds that a belief is justified if and only if the believer can't be properly blamed for violating any epistemic duty in holding the belief. Sosa accepts that in some sense it is good to be “justified” in this way. Yet surely there is more to internal epistemic excellence

than being blameless. After all, we might be blameless because we had been “brainwashed” or compelled by forces entirely outside of our control. We might be blameless despite being “deeply internally flawed” (Sosa 2003a: 159, 164). But our twins are not internally flawed. And any sort of “justification” for which “brainwashing” might suffice “is not of traditional epistemological concern, nor can it be the sort of epistemic rational state that we seek through inquiry into the rational status of our beliefs about the external world” (Sosa 2003a: 220).

Second, ontological internalism might be supported on the grounds that a belief is justified if and only if the believer accepts that the belief is sufficiently supported by the balance of evidence (or required by epistemic duty, or some such thing). Again Sosa accepts that in some sense it’s good to be “justified” in this way, but denies that it fully captures internal justification. For if a belief is to be justified in this way, then the believer presumably must also be *justified* in accepting that the belief is sufficiently supported by the balance of evidence. An unjustified acceptance won’t do. Yet if we add to the proposal that the acceptance is justified, then the proposal seems guilty of vicious circularity: invokes justification in characterizing justification (Sosa 2003a: 148, cf. 220–1). Moreover, such a view seems to violate the epistemic supervenience thesis.

Third, ontological internalism might be supported on the grounds that a belief is justified if and only if the believer would, upon the deepest and most sustained reflection, approve of holding it. Again Sosa accepts that this sort of “justification” is good in a

way, but denies that it fully explains the internal justification our twins enjoy. Even someone with irredeemably irrational fundamental commitments could be “justified” in the present sense (Sosa 2003a: 163–4). But our twins are not irrational at all.

With ontological internalism’s fortunes looking bleak, Sosa invokes virtue theory for an adequate explanation of the internal justification our twins enjoy. Earlier we noted that Sosa defines an intellectual virtue as a disposition to reliably believe the truth and avoid believing falsehoods. This is an incomplete specification. To better understand Sosa’s view, we must delve a bit deeper into the nature of dispositions.

Three points are especially important. First, dispositions are relative to an environment. I might be disposed to help a stranger if approached in broad daylight in a public space, but disposed to avoid that same stranger if approached in an alleyway at midnight. A bowling ball is disposed to roll when placed at the apex of a smooth steep hill, but disposed to remain stationary when placed at the nadir of the valley below. Second, an object’s dispositions are grounded in its intrinsic properties or “inner nature.” A bowling ball’s disposition to roll down a hill is grounded in its shape, texture and rigidity, properties that any molecular duplicate of the bowling ball would share. A similar point holds for a believer’s cognitive dispositions. Our cognitive disposition to form, or refrain from forming, a belief in certain conditions is grounded in the intrinsic properties of our minds, an inner nature that any mental duplicate of ours would share. Third, if two objects perfectly resemble one an-

other in their intrinsic properties, if they have the same inner nature, then they must have all the same dispositions relative to any environment.<sup>9</sup>

By now it should be obvious how Sosa proposes to handle the new evil demon thought experiment, and in particular how he proposes to explain the justification that our victimized twins enjoy. His basic proposal is that our twins are internally justified because they are intellectually virtuous. They are intellectually virtuous because of their “inner nature.” The inner, intrinsic quality of their minds is the same as ours, and so they are our equals in this respect. But — and this is the crux of the matter — ontological internalism is incapable of explaining what makes our inner nature virtuous: it is an incomplete view that must be supplemented by externalist virtue theory in a full accounting of internal justification. Our inner nature makes us virtuous because it suits us to perform well intellectually relative to an environment. And the fact that we are suited to perform well relative to an environment inevitably involves non-psychological facts about the environment. The same inner nature doesn’t suit us to perform well in just any environment, especially those populated by powerful, malevolent forces bent on deceiving us.

According to Sosa, when we judge that someone is justified in

---

<sup>9</sup> A fourth important point is that dispositions are relative to an overall internal condition. You might be disposed to remain calm when well-rested, but disposed to grow irritated when sleep-deprived. A bowling ball is disposed to roll down a hill when its surface is at roughly room temperature, but it isn’t disposed to roll when it’s so hot as to melt or deform on contact. For present purposes, I set aside this further detail of Sosa’s view.

believing something, we are judging that their belief is acquired through the exercise of one or more intellectual virtues, understood as truth-reliable cognitive dispositions. But dispositions and their reliability are relative to an environment. So when we judge that someone is justified in believing something, we are, at least implicitly, relativizing to an environment. Unsurprisingly, by default we relativize to what is a normal environment for us: a “normal human environment” (Sosa 1991: 143). Often such relativization occurs automatically “through contextual features not present to . . . consciousness” (2003a: 158). It might take considerable philosophical reflection to realize that this is what we’re doing.

Sosa’s claim that by default we evaluate our twins’ performance relative to a normal human environment receives support from experimental cognitive psychology. The new evil demon thought experiment primes us to think comparatively, comparing us and our twins. When humans are primed to think comparatively, they readily engage in what cognitive psychologists call “information transfer.” Information transfer occurs when judges rely on a “comparison standard” about which “they have abundant information available and which they have frequently used in the past” in order to simplify judgments about unfamiliar items. Instead of seeking information “about a judgmental target that they know very little about,” humans rely on “the rich and readily accessible information” encoded in the comparison standard (Mussweiler and Posten 2011: 1–2). This fits nicely with Sosa’s description of how we evaluate those peculiar victims of the fanciful malevolent demon: we

evaluative their performance relative to our normal human environment. It would be surprising if we did otherwise. Interestingly, the same body of psychological research suggests that comparative thinking induces humans to feel more certain in their judgments, and inclines them to bet more that they're right (Mussweiler and Posten 2011: 4). This helps explain the prevalence and resilience of favorable intuitive judgments about evil demon victims.

Here is how Sosa encapsulates his virtue-theoretic approach to justification, which has remained remarkably stable over the past twenty-five years, even if it has received increasingly sophisticated expression lately.

My proposal is that justification is relative to environment. Relative to our actual environment *A*, our automatic experience-belief mechanisms count as virtues that yield much truth and justification. Of course relative to the demonic environment *D* such mechanisms are not virtuous and yield neither truth nor justification. It follows that relative to *D* the demon's victims are not justified, and yet *relative to A their beliefs are justified*. Thus may we fit our surface intuitions about such victims: that they lack knowledge but not justification. (Sosa 1991: 144).<sup>10</sup>

Despite all that, there is for Sosa an important dimension of epistemic excellence along which we do outperform our victimized

---

<sup>10</sup> Compare Sosa 2003a: 156–161, and also Sosa 2009: 71–4, where he writes: “An important concept of justification involves evaluation of the subject as someone separable from her current environment. . . . [W]e might still enjoy such (internal) justification even when victims of the evil demon. . . . After all, the basis for evaluation is not the demon world but the actual world inhabited by the evaluators who are considering, as a hypothetical case, the case of the victim.”

twins. For although we and our twins are both equally virtuous relative to a normal human environment, our twins are not virtuous relative to the environment where their beliefs are actually formed, whereas we are virtuous relative to the environment where our beliefs are actually formed. This certainly seems to make our beliefs epistemically better than our twins' beliefs. Sosa has often called this sort of epistemic excellence "justification" (2003a: ch. 9; 2009: 192), but he has also shown a willingness to relinquish that terminology if it interferes with a proper appreciation of the status it denotes (e.g. Sosa 1991: 144, 289).<sup>11</sup>

Thus far we've focused on Sosa's engagement with ontological internalism. Now let's turn to his engagement with access internalism.

Access internalism is demanding and exceptionless: all factors that help determine a belief's justification must be unproblematically accessible to the believer from the armchair, via introspection and a priori insight. Reflectively inaccessible factors can't possibly make a difference, according to this view. Sosa rejects this on the grounds that there are clear counterexamples. Here are two:

Mary and Jane both arrive at a conclusion C, Mary through a brilliant proof, Jane through a tissue of fallacies. Each has now forgot-

---

<sup>11</sup> For punctilious readers dutifully checking the original sources, note that Sosa's earlier stipulative definitions of the terms 'apt' and 'adroit' differ importantly from his later stipulative definitions of those same terms. For example, compare Sosa 1991: 144, 289 and Sosa 2003a: ch. 9 to Sosa 2007: chs. 2 and 5. In the present chapter, I have chosen to restrict 'apt' and 'adroit' to their official meaning in Sosa's current system, where they name crucial statuses in the AAA-model of performance assessment, discussed in section 3 below.

ten much of her reasoning, however, and each takes herself to have established the conclusion validly. What is more, each of their performances is uncharacteristic, Jane being normally the better logician, Mary a normally competent but undistinguished thinker, as they both well know. The point is this: Jane would seem currently only better justified in taking herself to have proved C, as compared with Mary. As of the present moment, [given what each woman has access to from her armchair], Jane might seem as well justified as is Mary in believing C. We know the respective aetiologies, however; what do *we* say? Would we not judge Jane's belief unjustified since based essentially on fallacies? If so, then a belief's aetiology can make a difference to its justification. (Sosa 2003a: 151)

You remember having oatmeal for breakfast, because you did experience having it, and have retained that bit of information through your excellent memory. Your counterpart self-attributes having had oatmeal for breakfast, and may self-attribute remembering that he did so (as presumably you do), but his beliefs are radically wide of the mark, as are an army of affiliated beliefs, since your counterpart was created just a moment ago, complete with all of those beliefs and relevant current experiences. Are you two on a par in respect of epistemic justification? (Sosa 2003a: 152)

These cases demonstrate, Sosa claims, that it's possible for reflectively inaccessible factors to make a dramatic difference to justification. Mary is better justified in her belief than Jane, and your belief

is better justified than your twin's.<sup>12</sup>

Although Sosa rejects access internalism as a general theory of justification, he thinks that access internalists are on to something important. In this spirit, he proposes that there is a level of justification that does have an access requirement. Sosa calls this level of justification “reflective justification” and contrasts it with “unreflective justification,” which he often calls “animal justification” (1991: 291; 2003a: 228; 2009: 238–9). This brings bi-level epistemology into the picture front and center, though the virtue theory still remains center stage also, as we shall see.

Your belief that P is *unreflectively justified* just in case it is virtuously formed — that is, has its source in an intellectual virtue, unaided by reflection on your cognitive powers or circumstances. Your belief that P is *reflectively justified* just in case you are justified in believing that it is virtuously formed. Reflective justification involves developing, to a greater or lesser extent, a coherent “endorsing perspective” on your cognitive dispositions and environmental placement, which together determine how well justified your first-order beliefs are. From this endorsing perspective, you affirm that your basic ways of forming beliefs are reliable and virtuous, and form opinions about how your various first-order beliefs are justified due to their virtuous and reliable source. Reflective justification comes in degrees: the more coherent and detailed the perspective, the better reflectively justified you are in your relevant first-order beliefs.

---

<sup>12</sup> Greco 2005 develops this anti-externalist line of thought systematically.

Thus it is that Sosa imposes an access requirement on reflective justification. Reflective justification for your first-order belief that P requires you to have in view the factors that make your first-order belief unreflectively justified. Factors that are entirely hidden from you don't contribute to the reflective justification of your first-order belief, though they can contribute to its unreflective justification. It is critical to note, however, that Sosa does not restrict us to the armchair when accessing these epistemically relevant factors. Whereas traditional access internalists would chain us to the armchair, Sosa would liberate us, allowing perception, testimony and all manner of inquiry, both a priori and empirical, to inform our perspective and augment our access to relevant facts (Sosa 2009: 151). The armchair has its virtues and a role to play, but it's only a small part of a much larger repertoire at our disposal.

Just as unreflective justification must be produced by intellectual virtues, so too must reflective justification, in particular higher-order rational virtues involving self-awareness and critical reflection. Reflective justification combines virtue *and* perspective. We couldn't attain reflective justification without lots of antecedently acquired justified first-order beliefs, which provide the information needed to build up a view of our cognitive powers and the relevant features of our environment. These first-order beliefs are themselves acquired by first-order virtues, and are justified thereby, without any need for explicit reflective endorsement.

Must we also have a perspective on the operation and virtuosity of our higher-order virtues in order for them to do their work in

generating reflective justification? No, Sosa answers, further ascent isn't required. The fact that the perspective is virtuously produced and coherent is enough.

It would be absurd to require at *every* level that one must ascend to the next higher level in search of justification, and it seems equally absurd to suppose that a [meta-belief] can help justify an [object-level] belief, even though [the meta-belief] is itself unjustified . . . . The solution is to require the . . . coherence of a body of beliefs for the justification of its members, a coherence comprehensive enough to include meta-beliefs concerning object-level beliefs and the faculties [i.e. virtues] that give rise to them and the reliability of these faculties; but to allow that, at some level of ascent, justification is acquired by a belief as a belief that is non-accidentally true because of its virtuous source, and through its place in such an interlocking system of beliefs, without any requirement that it in turn must be the object of higher-yet beliefs directed upon it. (Sosa 1991: 293)

Charges of vicious circularity typically arise at this point, often accompanied by complaints that it is peculiarly dissatisfying that reflective justification could arise from the *mere fact* that beliefs are virtuously produced and coherently endorsed.<sup>13</sup> This raises the question of whether Sosa really can have his externalist cake and eat it too — whether he really can retain his commitment to externalism while at the same time “doing justice to the appeal of internalist intuitions.” Sosa's response to these matters takes us beyond the present chapter's scope, directly into the deep waters of the Prob-

---

<sup>13</sup> Sosa 2009 takes up the charges and complaints at great length.

lem of the Criterion and the Pyrrhonian Problematic. John Greco insightfully picks up the thread of Sosa epistemology at this point in chapter [ ] of this volume.

While a detailed accounting of the point falls beyond the scope of this chapter, it's worth noting that the two levels of justification that Sosa hypothesizes map nicely on to the standard view in contemporary cognitive science about how human cognition actually works. Sosa hypothesizes two levels or modes of human thought, one unreflective and mostly automatic, the other reflective and allied with deliberative agency. The unreflective level "is largely dependent on cognitive modules and their deliverances," and it is valuable that we are constituted to reliably and mostly automatically detect important truths. The reflective level monitors for the proper operation of the first-order modules and environmental influences, and strikes a balance when modular deliverances conflict or upset expectations. Such reflection is valuable not only because it can improve reliability by subjecting our "instinctive" doxastic habits to correction and "fine-tuning" (Sosa 2009: 142) but also because it enables "agency, control of conduct by the whole person, not just by peripheral modules" (Sosa 2004: 291–2); it allows us to "take charge . . . as a deliberative rational agent" (Sosa 2009: 138). Now compare all that to Daniel Kahneman's depiction of human thinking as involving two systems, what he calls "System 1" and "System 2."

System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.

System 2 allocates attention to the effortful mental activities that demand it, include complex computations. The operations of System 2 are often associated with subjective experience of agency, choice, and concentration. (Kahneman 2011: 20–1)

System 2 is slower and more cumbersome than System 1, but one thing it is good for is to help us “learn to recognize situations in which mistakes [on the first level] are likely and try harder to avoid significant mistakes when the stakes are high” (Kahneman 2011: 28).

### 3. *Knowledge, performance and safety*

The fundamental idea behind Sosa's theory of knowledge has remained essentially intact from at least the mid-1980s. All along he has maintained that knowledge is true belief “deriving from” or “out of” intellectual virtue (1991: 145, 277, *et. passim*). But beginning in the early 2000s, Sosa made a significant advance in how he formulated this definitive idea (beginning most conspicuously with Sosa 2003b). He developed an elegant general model of performance assessment, the AAA-model, and showed how his virtue-theoretic account of justification and knowledge is just an application of the general model. This new formulation is elegant and memorable, and consequently rhetorically effective. But it was no mere rhetorical improvement, however, because it makes evident previously unappreciated strengths and resources of the approach, and it even led to at least one noteworthy change in his definition of knowledge.

The AAA-model is simple and intuitive. We can assess perform-

ances for *accuracy*, *adroitness* and *aptness*. Accurate performances achieve their aim, adroit performances manifest competence, and apt performances are accurate because adroit. The-model applies to all conduct and performances with an aim, whether intentional, as in ballet dancing, or unintentional, as with a heartbeat.

Here is how the model applies in epistemology. Belief-formation is a psychological performance with an aim. For beliefs, Sosa identifies accuracy with truth, adroitness with manifesting intellectual virtue or — in the terminology Sosa has increasingly preferred — intellectual *competence*, and aptness with being “true because competent.” Apt belief, then, is belief that is true because competent. A competence in turn is “a disposition, one with a basis resident in the competent agent, one that would in appropriately normal conditions ensure (or make highly likely) the success of any relevant performance issued by it” (Sosa 2007: 29). Sosa identifies knowledge with apt belief.<sup>14</sup>

This approach to knowledge has three noteworthy benefits. First, it helps explain the added value of knowledge over mere true belief, an issue central to epistemology ever since Plato’s *Meno*. Succeeding through competence is better than succeeding by luck. A mere true belief could be had by luck, but not knowledge, which requires succeeding through competence (Sosa 2003; 2007: ch. 4; 2011: ch. 1). Second, as already mentioned, it places epistemic eval-

---

<sup>14</sup> A wrinkle added as of late: “A belief . . . might well be apt without being knowledge. Beliefs are *relevantly* apt only if they are believings *in the endeavor to attain truth*. This must now be understood implicitly in the account of animal knowledge as apt belief. The aptness of the belief must be in the endeavor to attain truth” (Sosa 2011: 21).

uation in a familiar pattern. Whether it's art, athletics, oratory or inquiry, we're keen to assess how outcomes relate to the relevant skills and abilities. The basic model of performance assessment applies across the entire range of evaluable rational activity: knowledge and epistemic normativity take their place as "just a special case" in this larger pattern (Sosa 2011: ch. 1). Third, it offers a solution to the Gettier problem. In a Gettier case, the subject believes the truth, and believes out of competence, but his belief isn't true because competent (Sosa 2007: 95–97).<sup>15</sup>

One noteworthy recent change in Sosa's view, prompted by the emergence of the AAA-model, is the abandonment of safety as a purported necessary condition on knowledge.<sup>16</sup> Previously Sosa claimed that knowledge requires belief that is both virtuous and safe (Sosa 1999, Sosa 2003a: 138–9). A virtuously formed belief is to be understood along the lines of unreflective justification discussed in the last section. A safe belief is one that is true and wouldn't easily have turned out false, at least not when it was formed on the same basis and through the same cognitive dispositions. The AAA-model subverts the safety requirement because a performance could be apt without also being safe. Indeed, it turns out that a performance can be apt despite being extremely unsafe.

Consider the performance of an archer who hits a bullseye because she shoots competently. Her shot is apt and the bullseye creditable to her. But consistent with that, her shot could have been un-

---

<sup>15</sup> See Turri 2011 for more on this solution to the Gettier problem.

<sup>16</sup> For much more on safety in Sosa's work, see Juan Comesaña's discussion in chapter [ ] of this volume.

safe: she might easily have missed. For example, she might have luckily avoided being drugged before the competition, which would have impaired her competence and resulted in a wild miss. Or a strong gust of wind, which would have ruined her shot, might have just been avoided by a rare confluence of local meteorological conditions. Despite performing aptly, she might still be in grave danger of failing in either of these ways: either through a serious threat to her competence or overall internal condition, or through a serious threat to the environment's normalcy and hospitality to her performance. But so long as the relevant relationship between the success and her competence remains, her performance remains apt and the bullseye remains creditable to her.

Given that Sosa identifies knowledge with apt belief, and given that aptness doesn't require safety, Sosa concludes that knowledge doesn't require safety either (2007: 28–9).

One principal consequence of abandoning safety is that it provides a new way of responding to *dream skepticism*. Evil demons and their doxastic victims are the stuff of philosophical fiction, but dreams are real and ubiquitous. Many of us have had the misfortune to occasionally mistake a dream for reality. Descartes worried that he might just be dreaming that he's seated near the fire. Does the real, acknowledged possibility that we might just be dreaming threaten our ordinary, waking perceptual knowledge? Can we really know based on sense experience if we might easily have been misled into believing the very same thing based on a dream that mimicked those sense experiences? The dream possibil-

ity is a much “closer” skeptical possibility than the demon world. And we might worry that its proximity renders our waking perceptual beliefs unsafe: too easily might we have been wrong, thanks to the ubiquity of dreams. In response, Sosa points out that this line of thought presupposes that knowledge requires safety. Having already rejected the safety condition on independent, general grounds, Sosa is perfectly positioned to defuse this line of skeptical reasoning (2007: chs. 2 and 5).

It's important to emphasize that giving up on safety as a requirement of knowledge does not require giving up on reliability as a requirement of competence. That is, abandoning safety doesn't mean abandoning reliabilism, which has long been front and center in Sosa's approach. On Sosa's view, in order to have a competence fit to produce apt shots, our archer must be reliably accurate in an environment normal for the practice of human archery. This is guaranteed by the definition, quoted above, of what counts as a competence: a disposition is a competence only if it “would in appropriately normal conditions ensure (or make highly likely) the success of any relevant performance issued by it.” But, as we saw earlier in the discussion of our victimized twins, the reliability and virtuosity of a disposition is relative to an environment. A disposition is virtuous because of what it enables us to accomplish in a normal environment. This approach neither prevents that same disposition from operating in other environments, even hostile ones, nor prevents it from producing in those other environments the same sort of success that it reliably produces in a normal environment.

When the right relationship between a reliable disposition and success obtains, the performance is apt and the outcome creditable to the agent.

In my view, abandoning safety brings Sosa's current view back in line with his most promising original vision for virtue epistemology. The addition of safety in the interim was an aberration. I say this for three reasons. First, the safety condition was motivated not as a way of clarifying or enhancing the basic virtue-theoretic approach, but rather by dialectical considerations, especially vis-à-vis the development of linguistic contextualist treatments of 'knows' that were influenced by Nozick's tracking theory of knowledge (Sosa 1999). Second, work done by an independent safety condition can equally be done by the virtue-theoretic apparatus, most centrally the aptness condition, so safety is superfluous, as can be gleaned from Sosa's own recent work (esp. 2007: chs. 2 and 5). Third, Sosa's recent explanation of why aptness doesn't require safety echoes features of his early explanation of what it is to believe out of intellectual virtue. For example, compare the "two interesting ways in which" a performance might be apt though unsafe, explained in Sosa 2007, to the two "interestingly different points" at which "things might have gone wrong" in belief formation, explained in Sosa 1991.

There are at least two interesting ways in which that shot might fail to be safe . . . . The following two things might each have been fragile enough to deprive the shot of safety: (a) the archer's level of competence, for one, and (b) the appropriateness of the condi-

tions, for another. Thus (a) the archer might have recently ingested a drug . . . so that his blood content of the drug might too easily have slightly higher, so as to reduce his competence to where he would surely have missed. Or else (b) a freak set of meteorological conditions might have gathered in a way that too easily a gust might have diverted the arrow on its way to the target. (Sosa 2007: 28)

If S believes a proposition in field F, about the shape of a facing surface before him, . . . things might have gone wrong at interestingly different points. Thus the medium might have gone wrong unknown to the subject, and perhaps even unknowably to the subject; or something within the subject might have changed significantly: thus the lenses in the eyes of the subject might have become distorted, or the optic nerve might have become defective in ways important to shape recognition. If what goes wrong lies in the environment, that might prevent the subject from knowing what he believes, even if his belief were true, but there is a sense in which the subject would remain subjectively justified or anyhow virtuous in so believing. (Sosa 1991: 139–40)

The second quote strongly suggests that if neither of those things does go wrong, then the subject believes virtuously, with no hint that a true belief thus formed couldn't be knowledge. A hostile environment *might* prevent a virtuously formed true belief from counting as knowledge. Not 'must' or even 'would', but 'might'.

#### 4. *Meta-aptness and knowing full well*

In addition to unreflective or "animal" knowledge ("apt belief") and

reflective knowledge (“apt belief aptly noted”), Sosa has recently added another noteworthy epistemic category: knowing full well. He does this by importing another dimension of performance assessment: meta-aptness. A performance is apt, as already mentioned, just in case it is accurate because adroit. A perform is meta-apt just in case it is “well-selected” (2011: 8). Selecting well is a matter of competent risk management, and this requires a perspective on your abilities and relevant environmental factors that influence your likelihood of succeeding.

Sosa illustrates this additional dimension by building on the archery example used to illustrate the AAA-model initially.

Let our archer now be a hunter rather than a competitor athlete. Once it is his turn, the competitor must shoot, with no relevant choice. True, he might have avoided the competition altogether, but once in it, no relevant shot selection is allowed. The hunter by contrast needs to pick his shots, with whatever skill and care he can muster. Selecting targets of appropriate value is integral to hunting, and he would also normally need to pick his shots so as to secure a reasonable chance of success. The shot of a hunter can therefore be assessed in more respects than that of a competitor athlete. The hunter’s shot can be assessed twice over for what is manifest in it: not only in respect of its execution competence, but also in respect of the competence manifest in the target’s selection and in the pick of the shot. (Sosa 2011: 5–6)

One major benefit of acknowledging meta-aptness is that it allows us to appreciate what the hunter does right when he *forbears*, or chooses to not take a shot. The aim of hunting is to bring down

prey. But forbearing is no way to bring down prey, because by forbearing the hunter *automatically* fails thereby to bring down the prey. If the only aim were bringing down prey, we would have no way to positively assess the hunter's forbearance. Yet the hunter's meta-judgment that he should bide his time for a better opportunity seems clearly wise and suitable in many cases. The category of meta-aptness explains this further positive dimension of his performance.

A performance can be apt without being meta-apt. If the hunter decides to take what is, even by his own lights, an unwise shot, but overcomes the difficulties and brings down the prey through a skillful shot nonetheless, the shot is apt, successful because competent. But it isn't meta-apt, because the hunter should not have shot, even by his own lights. It is a foolhardy shot, one that doesn't derive from competent risk management. Similarly, a performance can be meta-apt without being apt. Conditions might be very conducive to success, which the hunter well appreciates, which in turn motivates him to release an adroit shot that, improbably, misses. The shot is inaccurate and so, by definition, inapt. But it's still meta-apt.

Once we have aptness and meta-aptness in view, we can then assess performances for how the two categories relate. A truly expert and rational performance is one where the performance is apt because meta-apt. Here we reach a new height of accomplishment.

Applying these insights to epistemology, we see how *knowing full well* fits in. You know full well that P just in case you aptly believe that P, and you aptly believe that P because you com-

petently assessed your propensity to believe the truth in the context where your belief was formed. The meta-competence involved here “governs whether or not one should form a belief at all on the question at issue, or should rather withhold belief altogether.” This requires having a perspective on your abilities and environmental conditions. In knowing full well, you reach new “epistemic heights” (Sosa 2011: 12).

## 5. *Conclusion*

Sosa’s bi-level virtue epistemology is wide-ranging, powerful and compelling. It casts fresh light on a host of fundamental questions in epistemology, teaching us much of value in the process. It has inspired an entire research program, contemporary virtue epistemology, and a veritable flood of sympathetic and critical responses, perhaps more than any other epistemological project in recent memory. And Sosa is still actively improving the view and applying it in new directions. I lack the space here to review the rich and growing literature surrounding Sosa’s work. I also lack the space to review Sosa’s impressive contributions to the history of epistemology, which identify numerous historical figures who endorsed, or at least had sympathy for, bi-level virtue theory themselves (see especially Sosa 2009: part I, and Baron Reed’s discussion in chapter [ ] of this volume). Instead, I’ve limited myself to reviewing the main topics that Sosa addresses with his bi-level virtue theory: the nature of epistemic justification and knowledge, and the allied topics of

epistemic normativity and skepticism. A judicious study of Sosa's considerable body of work will reveal that I've barely managed to do justice even to that.<sup>17</sup>

*Word count: 9526*

---

<sup>17</sup> For helpful conversation and feedback, I'm happy to thank Ian MacDonald, Ernest Sosa and Angelo Turri.

## References

- Conee, Earl and Richard Feldman. 2001. Internalism defended. Reprinted in *Evidentialism: essays in epistemology*. Oxford University Press, 2004.
- Dancy, Jonathan, Ernest Sosa and Matthias Steup, eds. 2010. *Companion to epistemology*, 2ed. Wiley-Blackwell.
- Greco, John. 2005. Justification is not internal. *Contemporary debates in epistemology*. Ed. Matthias Steup and Ernest Sosa. Blackwell.
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. Doubleday Canada.
- Mussweiler, Thomas and Ann-Christin Posten. 2011. Relatively certain! Comparative thinking reduces uncertainty. *Cognition* (2011), doi:10.1016/j.cognition.2011.10.005.
- Sosa, Ernest. 1991. *Knowledge in perspective*. Cambridge University Press.
- Sosa, Ernest. 1999. How to defeat opposition to Moore. *Philosophical perspectives* 13: 141–53.
- Sosa, Ernest. 2003a. Beyond internal foundations to external virtues. *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Blackwell.
- Sosa, Ernest. 2003b. The place of truth in epistemology. *Intellectual virtue: perspectives from ethics and epistemology*. Ed. Michael DePaul and Linda Zagzebski. Oxford University Press.
- Sosa, Ernest. 2007. *Apt belief and reflective knowledge, volume 1: a virtue epistemology*. Oxford University Press.
- Sosa, Ernest. 2009. *Apt belief and reflective knowledge, volume 2: reflective knowledge*. Oxford University Press.
- Sosa, Ernest. 2011. *Knowing full well*. Princeton University Press.
- Turri, John. 2009. On the general argument against internalism. *Synthese* 170.1: 147–53.
- Turri, John. 2010. Epistemic supervenience. *Companion to epistemology*, 2ed. Ed. Jonathan Dancy, Ernest Sosa and Matthias

Steup. Wiley-Blackwell.

Turri, John 2011. Manifest failure: the Gettier problem solved. *Philosophers' imprint* 11.8: 1–11.