An appraisal of the evolution of some of the main themes in Sosa's epistemology

Juan Comesaña [1]

Abstract: In this article, I propose to trace the evolution of three central concepts in Sosa's epistemology: the distinction between animal and reflective knowledge, closure principles, and the safety condition. These three planks played a central role in the early presentations of Sosa's epistemology, but have recently undergone interesting changes.

Keyword: Ernest Sosa. Virtue Epistemology. Safety. Closure. Reflection.

Introduction

Ernest Sosa is, without a doubt, one of the most important philosophers of the past fifty years. His contributions to epistemology, in particular, have had (and continue to have) a major impact in the development of that subdiscipline. On a more personal note, Ernie was my advisor at Brown in the late 90s and early 2000s. I had a very good time as a graduate student, and my relationship with Ernie played no small role in that. I arrived at Brown from Argentina with a solid formation in logic, philosophy of science and philosophy of language, but with a somewhat modest command of English, a very spotty and idiosyncratic formation in epistemology, and the understandable nervousness of starting a new life far from home. It goes without saying that from Ernie I learned epistemology, but it perhaps does not go without saying that he also helped me enormously in the more personal aspects of adapting to my new life, and for all that I will always be grateful. It is thus a pleasure for me to take part in this tribute to his philosophy.

[1] University of Arizona, Tucson – United States of America. https://orcid.org/0000-0002-5475-3759   E-mail: comesana@email.arizona.edu

I propose to trace the evolution of three central concepts in Sosa's epistemology: the distinction between animal and reflective knowledge, closure principles, and the safety condition. These three planks played a central role in the early presentations of Sosa's epistemology, but have recently undergone interesting changes.

1 Sosa on safety

As far as I am aware, Sosa first talked about safety as a condition on knowledge in Sosa (1996), and devoted a full paper to it in Sosa (1999). He considerably modified his attitude towards safety in Sosa (2007), where he no longer considers safety as a straightforward condition on knowledge, and that modified attitude continues to this day in Sosa (2021).

Sosa's safety condition is best understood in the context of Nozick's sensitivity condition in Nozick (1981). According to Nozick's theory of knowledge presented in that book, and to a first approximation:

S knows that p if and only if:

p is true;

S believes that p;

If p were not true, then S would not believe p;

If p were true, S would still believe it.

Condition 3 is Nozick's sensitivity condition. That condition allows for knowledge of ordinary propositions, such as the proposition that I have hands, but it blocks knowledge of the negation of skeptical scenarios, such as the hypothesis that I am a handless brain in a vat. Moreover, due to this behavior of the sensitivity condition, Nozick's theory of knowledge has it that knowledge is not closed under known logical implication (i.e., it allows that a subject can know a proposition p without knowing a proposition q which the subject knows is entailed by p).

Nozick's sensitivity condition explains the intuition many have that we do not know that skeptical scenarios do not obtain– it explains this by entailing it. Sosa agrees that there is this strong intuition, but he points out that it is counterbalanced by an equally strong intuition on the part of many others

that we do of course know that we are not the victims in skeptical scenarios. How could we account for *both* of these intuitions? According to Sosa, we can do that by requiring that knowledge be *safe* rather than sensitive. Sosa formulated his safety condition in several ways, but the following formulation will do for our purposes:

> **Sosa's safety condition**: *S* knows that *p* only if, if S were to believe that *p*, then *p* would be true.

Notice that Sosa's safety condition is the contraposition of Nozick's sensitivity condition.

However, the conditions are not equivalent if, as it is usually assumed, subjunctive conditionals do not contrapose. We can see this by focusing precisely on the proposition that skeptical scenarios do not obtain. Our belief in that proposition is not sensitive because, if skeptical scenarios were to obtain, we would still believe that they do not (this is part of what makes skeptical scenarios so insidious). However, our belief that skeptical scenarios do not obtain is safe, because if we were to believe it (even under slightly different circumstances), it would still be true. Of course, this explanation of the difference between Nozick's sensitivity and Sosa's safety rests on the claim that we are not in fact the victims in a skeptical scenario, but Nozick himself already presupposes that that is true (even though we do not know it).

How does Sosa's safety condition, then, explain the two rival intuitions regarding whether we know that skeptical scenarios do not obtain? As follows: we do know that skeptical scenarios do not obtain, because our belief that skeptical scenarios do not obtain is safe (and the other conditions for knowledge are also satisfied), but we naturally think that we do not know it because safety is easily confused with sensitivity (due to the fact that they are contrapositives of each other). Our beliefs in the negation of skeptical scenarios are indeed safe, but safety is easily confused with sensitivity, and our belief in the negations of skeptical scenarios are not sensitive. Moreover, safety but not sensitivity is required for knowledge. As a result, although we do know that skeptical scenarios do not obtain, it is understandable why we would think that we do not know it.

But is safety indeed a condition on knowledge? Many, including myself, have argued that it is not. My own counterexample in Comesaña (2005, p. 397) was the following:

> HALLOWEEN PARTY: There is a Halloween party at Andy's house, and I am invited. Andy's house is very difficult to find, so he hires Judy to stand at a crossroads and direct people towards the house (Judy's job is to tell people that the party is at the house down the left road). Unbeknownst to me, Andy doesn't want Michael to go to the party, so he also tells Judy that if she sees Michael she should tell him the same thing she tells everybody else (that the party is at the house down the left road), but she should immediately phone Andy so that the party can be moved to Adam's house, which is down the right road. I seriously consider disguising myself as Michael, but at the last moment, I don't. When I get to the crossroads, I ask Judy where the party is, and she tells me that it is down the left road.

If you agree with me that I do know that the party is down the left road, then that is in principle a problem for safety as a condition on knowledge, for it looks as if that belief of mine is not safe.

Sosa himself later changed his mind about safety. In Sosa (2007), Sosa synthesized his virtue epistemology in a new framework. A true belief is an *accurate* belief; a belief which is the result of a competence housed in the subject is an *adroit* belief; and a belief which is accurate *because* it is adroit is an *apt* belief. In order to amount to animal knowledge, a belief must be apt. Now, does aptness require safety? No, Sosa now answers, appealing to his own counterexample:

> **Jokester**: You see a surface that looks red in ostensibly normal conditions. But it is a kaleidoscope surface controlled by a jokester who also controls the ambient light, and might as easily have presented you with a red-light + white-surface combination as with the actual white-light + red-surface combination. Do you then know the surface you see to be red when he presents you with that good combination, despite the fact that, even more easily, he might have presented you with the bad combination? (SOSA, 2007, p. 31).

Sosa's answer to his own question is that you do have animal knowledge that the surface is red, for your belief is apt– i.e., accurate because adroit– despite the close possibility of error. Now, Sosa goes on to argue that the close possibility of error does rob you of *reflective* knowledge– i.e., you do not know

that you know that the surface is red. In this way, safety still plays a special role for the Sosa of 2007– it is no longer considered a condition on animal knowledge, but it is still considered a condition on reflective knowledge.

Moreover, safety still plays a second role for the 2007 version of Sosa's epistemology. For, despite the fact that a belief can be apt, and this amounts to animal knowledge, even if it is not safety, safety itself enters into the specification of what it is for a belief to be apt. Sosa says:

> **Aptness**: For any correct belief that p, the correctness of that belief is attributable to a competence only if it derives from the exercise of that competence in appropriate conditions for its exercise, and that exercise in those conditions would not then too easily have issued a false belief. (SOSA, 2007, p. 33).

For a belief to be apt, remember, it must be accurate *and* its accuracy must be explained by its adroitness. To say that a belief is adroit is to say that it derives from a competence the subject has. A competence is a disposition to produce beliefs under certain conditions, but not all such dispositions are a competence. For a belief-forming disposition to be a competence, its exercise under the appropriate conditions must result in safe beliefs. Thus, take the disposition to believe that there is a snowball in front of you when you have an experience as of a snowball in front of you. That disposition is a competence of yours just in case, whenever you have an experience as of a snowball in front of you and the conditions are the appropriate ones for the exercise of that disposition, not easily it could have happened that you had that same experience without its being the case that there really was a snowball in front of you. This gives rise to what I once called "competence safety":

> **Competence safety**: If S knows that p in virtue of exercising a disposition to believe under conditions that are normal for the exercise of that disposition, then not easily would S believe that p in virtue of exercising that competence under those conditions without its being so that p. (COMESAÑA, 2013, p. 9).

A crucial question now is: what determines what are the normal conditions for the exercise of a belief-forming disposition? At the time, at least, Sosa seemed happy to grant that the appropriate conditions for the exercise of a competence are those such that, were the competence to be exercised under those conditions, it would yield beliefs that amount to knowledge because it

would yield beliefs that are safe. If there are no such conditions for a given disposition, then that disposition does not constitute a competence.

I believe that what is essentially the same picture about the role of safety in epistemology is still present in Sosa's most recent book, *Epistemic Explanations*. There, Sosa holds that while safety is not necessary for animal knowledge, it is necessary for "reflective knowledge full well", for this kind of knowledge requires not only aptness, but also aptly attained aptness, and this can only be achieved if one is safely exercising one's competence.

Let's recap. In the late 90s, Sosa proposed the safety condition as a better modal condition on knowledge than Nozick's sensitivity condition. By the late 2000s, however, the place of safety in Sosa's epistemology was considerably subtler, and it remains that way up to this day. First of all, Sosa holds that animal knowledge does not require safe belief. Second, he holds that safety is still necessary for reflective, full well knowledge, and third, he holds that what determines whether a belief-forming disposition is a competence– and, thus, a possible source of adroit beliefs– is whether there are conditions under which such a disposition produces safe beliefs.

Sosa's current position on safety is an unstable one, I believe. The reasons for thinking that animal knowledge is possible even in the absence of safe belief seem to be equally good reasons for thinking that reflective knowledge is also possible without safe belief– after all, reflective knowledge is for Sosa, at this point, second order knowledge, and if close possibilities of error are not incompatible with first-order knowledge, why would they be incompatible with second-order knowledge? Now, there is an answer to that question, but it is not one that Sosa would be happy with. The answer I'm thinking of appeals to the Williamsonian idea of margins for error in knowledge. Williamson's idea, related to the notion of safety, is that our fallibility entails that knowledge requires a margin for error– for instance, there is a limit to the resolution to which we can know by sight the height of a tree. We can know, for instance, that it is between 9 and 10 feet, but we cannot know just by looking at it that it is exactly 9 feet, 5, and ¾ inches. So far, so uncontroversial. Williamson assumes, however, that the margin for error moves with the distance from the real height. In other words, we can know that the tree is between 9 and 10 feet if it is towards the middle of that range, but not if it is towards one of the extremes. This conception of how safety works does entail that higher levels of knowledge require tighter margins of error, and thus does have the consequence that iterated knowledge

is harder to have. But an alternative conception is that if the tree is between 9 and 10 feet and I believe this out of a competence of mine, then I know that it is between 9 and 10 feet. I believe that Sosa would be friendlier to this alternative conception than to Williamson's. But this cuts off the obvious route to holding that the very unsafety that is compatible with first-order knowledge prevents us from having second-order (or reflective) knowledge.

With respect to the role of safety in the definition of a competence, that too seems like an unstable position. A disposition amounts to a competence, remember, just in case there are conditions under which it would result in safe beliefs. But this seems to count *almost any* belief-forming disposition as a competence, because, for almost any given belief-forming disposition, there will be *some* conditions in which it produces safe beliefs. Take what we would normally consider a perverse belief-forming disposition, such as the disposition to believe that there is a snowball to one's left when one has an experience as if there is a snowball in front of one. In our circumstances, this disposition of course does not produce safe beliefs, but there are possible circumstances under which it does: for instance, environments where the light behaves in such a way that whenever objects appear to be in front of the perceiving subjects, they are actually to their left. This means that even such a perverse disposition will count as a competence, which seems clearly the wrong result (and, at any rate, not a result expected by Sosa).[2]

3 Sosa on closure

The discussion of closure principles has played a key role in contemporary epistemology (as well as in the interpretation of the modern epistemology of Descartes, for instance). According to a rough characterization of closure (the only one we will need), if a subject knows that p and knows that p entails q, then the subject knows (or is in a position to know) q.

Sosa used to be an advocate of closure, as can be gleaned from his appeal to safety. For remember that the appeal to safety was meant to replace Nozick's appeal to sensitivity. Nozick's account of knowledge has the consequence that knowledge is not closed, and Sosa used to think that one advantage of

---

[2] Why do I say that for *almost* any disposition there will be circumstances under which it produces safe beliefs? Because there are dispositions to believe propositions which are true under no circumstances (such as logical contradictions), or propositions which are guaranteed to be false when believed (such as the proposition that there are no believers), etc.

safety over sensitivity is that it does not require one to give up closure (see, for example, Sosa (2004)). However, in his most recent book Sosa abandons closure.

Sosa's recent rejection of closure is related to the analogy between judgment and agency which permeates his most recent expositions of virtue epistemology. Sosa holds that any action requires the assumption that the conditions for its exercise are appropriate. This is an automatic assumption, not based on any independent evidence. For that very reason, these assumptions cannot amount to knowledge. But what goes for action generally also goes for judgment: the exercise of the ability to judge also requires assuming that the conditions for the proper exercise of that ability are met, and this assumption is, again, automatic and not founded on any evidence. Thus, when I judge that there is a snowball in front of me I automatically assume that I am not a brain in a vat being fed the relevant experience in the absence of any nearby snowball, but I do not know that this assumption is true. Thus, Sosa seems to now agree with Nozick regarding the "solution" to the skeptical problem– namely, he now holds that although we do of course know ordinary propositions such as that there is a snowball in front of us, we do not know some of what is assumed in attributing that knowledge, namely, that we are not the victims of a skeptical scenario.

It is worth pointing out here that what seems to be driving Sosa to deny the principle of closure is at least very close to what Stew Cohen called the "problem of easy knowledge" (see, for instance, Cohen (2002)). In my version of the problem, it arises for any theory according to which it is possible to know some proposition p on the basis of some evidence E even without knowing that E is not in this case misleading. Given that p itself entails that E is not misleading evidence for p (that E is misleading evidence for p means that although E is evidence for p, E is false), the closure principle has it that in order to know p one must know that E is not misleading evidence for p. But how can one this? There seem to be only two options: either one knows it on the basis of p itself, or on the basis of E. But one cannot really know it *only* on the basis of p, for one knows p itself (if at all) on the basis of E. The only possible justification for the proposition that E is not misleading evidence for p, then, seems to be E itself. But notice that the proposition that E is misleading evidence for p (i.e., the proposition that whereas E is true p is false) entails E. How could we then accept that E is not misleading evidence for p on the basis of E itself? According to the proposition that E is not misleading

evidence for *p*, *E* is true, and so the truth of *E* cannot conspire against that proposition. Therefore, it would seem that, against the closure principle, it is not actually possible to know that *E* is not misleading evidence for *p*. [3]

We know what Sosa's answer would be to that problem: give up closure. But there is an alternative answer in the literature which fits well with Sosa's virtue epistemology. According to this proposal, there is a third alternative regarding what can justify us in believing that *E* is not misleading evidence for *p* (in addition to *p* and *E*): namely, we can be justified *a priori* in believing that proposition. Moreover, this *a priori* justification (according to the proposal in question) is based on our inferential competence (versions of this view have been proposed by Cohen (2010) and Wedgwood (2012)). For when we assume that *E* can justify us in believing *p*, we assume that we have the competence to (defeasibly) infer *p* from *E*. If we do in fact have that competence, then we can deploy it not only "online", when faced with *E* itself, but also "offline", when we merely *assume* that *E* is true. If we assume that *E* is true, then we can use our inferential competence to infer that *p* under the scope of that assumption, and so we can conclude (defeasibly) that if *E* is true, then *p* is also true. And this is precisely the proposition that *E* is not misleading evidence for *p*. Whenever we know that *p* on the basis of *E*, then, we have available an *a priori* justification for believing that *E* is not misleading evidence for *p* which is independent of both *E* and *p*.

Thus, Sosa's abandonment of closure seems premature in the face of this alternative solution to the problems which led to that abandonment, an alternative very friendly indeed to a virtue epistemology.

4 Sosa on reflective knowledge

We have already had occasion to recall Sosa's distinction between animal and reflective knowledge, which has been a staple of Sosa's virtue epistemology from the beginning. In Sosa (2015), Sosa appeals to the distinction between animal and reflective knowledge to explain our intuitive reaction to certain famous cases in the epistemological cannon, such as the fake barns case. In this case, which Carl Ginet suggested to Alvin Goldman (see Goldman (1976)), a subject is driving through an area filled with fake barns which cannot be distinguished by sight from real barns, but happens to fix his gaze on the only real barn around and believes that it is a barn. Several philosophers believe

---

[3] For more on this issue, see Comesaña (2020) and Comesaña and Sartorio (2014).

that the subject does not know that he is seeing a real barn, but several others suspect that he does know. Sosa creates several cases of his own on this style, including the case of Simone, a subject who, as part of her training as a pilot, uses a flight simulator which is indistinguishable from the real thing. Let us suppose that Simone graduates to flying real airplanes, but her trainers still put her, from time to time, in the simulator without telling her about it. The question then is: does Simone know that she is flying a real plane when she is? In Sosa (2015), Sosa replies to this question by saying that whereas Simone has animal knowledge, she lacks reflective knowledge.

Sosa thinks that Simone has animal knowledge because her belief is adroit: accurate because apt. It is true that the conditions for the exercise of Simone's competence to tell that she is flying a real plane are modally fragile, but all that means is that Simone's animal knowledge is not safe– and we already know that, by this point, Sosa has abandoned safety as a condition on knowledge. But that very modal instability prevents Simone from having reflective knowledge.

We said before, in our discussion of safety, that it was strange for Sosa to hold that the modal instability of the conditions for the proper application of a competence are not an obstacle to animal knowledge but do prevent reflective knowledge. Perhaps for reasons similar to these, in his most recent book (SOSA, 2021) Sosa now holds that Simone also has reflective knowledge. She still lacks something, however, and that is *secure* knowledge:

Secure Knowledge

A given judgment constitutes secure knowledge only if the thinker is safe from the following fate: *losing* their pertinent complete SSS competence to so judge while at the same time *retaining* a disposition to make judgments when they "inquire" into that question *even absent any such competence* (on a broad notion of inquiry requiring only that the thinker take up that question, at least implicitly and representationally, if not through intentional conscious questioning or affirmation).

(SOSA's "SSS" competence refers to the skill, shape and situation that must be in place for a competence to be appropriately exercised.) Thus, by definition, secure knowledge requires that the conditions for the exercise

of the competence obtain in a modally stable way. Therefore, even though Simone enjoys animal and even reflective knowledge that she is flying a real plane when she is, she does not have secure knowledge (because too easily she could have judged that she was flying a real plane while she was not).

Sosa is of course free to stipulate a notion of knowledge which requires the modal stability of the circumstances necessary for the exercise of the competence in question, but the substantive question is whether any such notion cuts epistemic nature at its joints. Once it is granted that Simone can have even reflective knowledge, what importance can it have that she does not have secure knowledge? Is this notion of securing knowledge tracking anything that we would pre-theoretically accord normative significance to?

A full answer to that question would require careful study of Sosa's latest book, a task which in any case is very worthwhile. Here I will simply point out something worth keeping in mind when undertaking that study. We should keep firmly in mind that there is an obvious risk that Simone is running, namely, the risk of being in a simulator. But it is not obvious that the risk in question has any normative counterpart. One possibility is that Simone satisfies all the relevant epistemic statuses, but does so *riskily*. Sosa's history of dealing with fake-barn-style cases suggests that he thinks that when subjects satisfy an epistemic status riskily, then there is another status which they do not satisfy, because it is impossible by definition to do so riskily. That status was at some point reflective knowledge, and it is now secure knowledge. But why is it not an option that *all* epistemic statuses worth talking about (namely, those that track something that we would independently ascribe normative importance to) can be satisfied riskily?

Conclusion

It is a pleasure for me to contribute to this volume in honor of Ernest Sosa. The fact that, in our profession, a common way of expressing admiration for authors is to criticize them should not obscure the fact that I do indeed admire Ernie very much. I count myself very lucky indeed to have had him as my advisor, and even more lucky to count him as a friend.

COMESAÑA, J. Uma avaliação da evolução de alguns dos principais temas da epistemologia da Sosa. *Trans/form/ação*, Marília, v. 44, p. 29-34, Edição Especial - Dossier "Ernest Sosa", 2021.

Resumo: Neste artigo proponho traçar a evolução de três conceitos centrais na epistemologia de Sosa: a distinção entre o conhecimento animal e reflexivo, os princípios de fechamento e a condição de segurança. Estes três temas desempenharam um papel central nas primeiras apresentações da epistemologia de Sosa, mas, recentemente, sofreram mudanças interessantes.

Palavras-chave: Ernest Sosa. Virtue Epistemology. Segurança. Fechamento. Reflexão.

REFERENCES

COHEN, S. Basic Knowledge and the Problem of Easy Knowledge. **Philosophy and Phenomenological Research,** 65 (2): 309–29, 2002.

COHEN, S. Bootsrapping, Defeasible Reasoning and A Priori Justification. **Philosophical Perspectives,** 24 (1), p. 141-159, 2010.

COMESAÑA, J. Unsafe Knowledge. **Synthese,** 146, p. 395–404, 2005.

COMESAÑA, J. Safety and Epistemic Frankfurt Cases. In: **Virtuous Thoughts:** The Philosophy of Ernest Sosa, edited by John Turri, 165–78. Springer. https://link-springer-com.ezproxy2.library.arizona.edu/chapter/10.1007/978-94-007-5934-3_9, 2013.

COMESAÑA, J. **Being Rational and Being Right**. Oxford: Oxford University Press, 2020.

COMESAÑA, J., AND SARTORIO, C. Difference-Making in Epistemology. **Noûs,** 48 (2), p. 368–87, 2014.

GOLDMAN, A. Discrimination and Perceptual Knowledge. **Journal of Philosophy,** LXXIII (20), p. 771–91, 1976.

NOZICK, Robert. **Philosophical Explanations**. Harvard: Harvard University Press, 1981.

SOSA, E. **Epistemic Explanations:** A Theory of Telic Normativity, and What It Explains. Oxford: Oxford University Press, 2021.

SOSA, E. Postcript to 'Proper Functionalism and Virtue Epistemology.' In: Jonathan Kvanvig. (Ed.): **Warrant in Contemporary Epistemology:** Essays in Honor of Plantinga's Theory of Knowledge. Lanham: Rowman and Littlefield, 1996.

SOSA, E. How to Defeat Opposition to Moore. **Philosophical Topics** 174, p. 1001–37, 1999.

SOSA, E. Relevant Alternatives, Conxtextualism Included. **Philosophical Studies,** 119 (1/2), p. 36–65, 2004.

SOSA, E. **A virtue epistemology**: Apt belief and reflective knowledge, vol 1. Oxford: Clarendon Press, 2007.

SOSA, E. **Judgment and Agency**. Oxford: Oxford University Press, 2015.

WEDGWOOD, R. A Priori Bootstrapping. In: CASULLO, Albert; THUROW, Joshua C. (Eds.) **The A Priori in Philoophy**. Oxford: Oxford University Press, 2012.

COMESAÑA, J.