A CRITICAL ASSESSMENT OF SOSA'S "TRANSCENDENTAL ARGUMENT" IN KNOWING FULL WELL[1]

CLAUDIO CORMIK
https://orcid.org/0000-0003-0162-2429
Sociedad Argentina de Analisis Filosofico
Department of Philosophy
Buenos Aires
Argentina
ccormik@ccaece.edu.ar

Article info
CDD: 121
Received: 26.12.2019; Revised: 09.03.2020; Accepted: 10.03.2020
https://doi.org/10.1590/0100-6045.2020.V43N1.CC

Keywords
Epistemology
Sosa, Ernest
Transcendental arguments
Naturalism

Abstract: In a provocative, yet scarcely discussed, argument at the end of Knowing Full Well, Ernest Sosa has attempted to determine what kind of evidence we possess in support of the belief that our cognitive capacities as human beings are reliable. According to Sosa, we can appeal to considerations of coherence to prove that

[1] I wish to thank an anonymous referee for *Manuscrito* for valuable remarks about a previous version of this article. I would also like to thank my colleagues in the epistemology group at SADAF (Santiago Armando, Jonathan Erenfryd, Anahí Grenikoff, Pedro Martínez Romagosa, Bruno Muntaabski, Daniel Pared, Federico Penelas, Alejandro Petrone, Moira Pérez, Blas Radi, Florencia Rimoldi and Mauro Santelli) to which this work was first presented.

such capacities are reliable (i.e., it would be epistemically self-defeating to think otherwise). However, Sosa also declares that such considerations are not "determinative, ultima facie" reasons—which is to say, they are to be regarded as defeasible. As we will try to point out, this overall strategy is ultimately incoherent. Furthermore, as we will argue, Sosa fails in attempting to provide us with an analogy between the case of doubting the reliability of the cognitive faculties of an individual and doubting such reliability in the case of the species.

1. OUTLINE

The present work will attempt to evaluate Sosa's "transcendental argument" concerning the reliability of human knowing faculties. We will proceed as follows: in *section 2*, we will analyze the way in which Sosa presents the alleged doubts that, in a naturalistic framework, would appear concerning our own reliability as knowing subjects. In *section 3*, we will introduce Sosa's proposed solution for these doubts, which appeals to an analogy between the case of doubting the capacities of the human species and doubting one's capacities as an individual. In *section 4*, we will tackle some specific aspects of the argument: the problem of why the skeptical attitude is supposed to be "incoherent"; what "faculties" the argument is considering; how the argument concerning the case of the species is supposed to be analogous to the case of the individual, and how the agent's own perspective on her cognitive capacities relates to her actual cognitive situation. This last element will lead us, in *section 5*, to present the way in which Sosa *weakens* his "transcendental argument": given that the argument, as it first appears, is "insensitive to the facts" concerning the subject's *actual* situation, Sosa will hold that such argument does not provide us with "determinative, *ultima facie*" reasons. In *section 6*, we will introduce our first criticism to

Sosa's position in this weakened form: as we will argue, the author cannot prove that the agent's own reasons can somehow be combined with information only available from a third-person perspective. Furthermore, as we will argue in *section 7*, things get even worse for Sosa's argument when we take seriously his *Disablex* argument *as an analogy* and consider its applicability to the capacities of the species as such: in this case, unlike the case of the capacities of the individual, we find that normal human capacities are simply the "standard meter" on the basis of which we determine what it is to be "reliable" capacities−and thus it is simply false that we need to commit to a certain contingent description of their origins to deem them as reliable.

2. SOSA, NATURALISM AND DOUBTS CONCERNING THE RELIABILITY OF HUMAN KNOWING CAPACITIES

In *Knowing Full Well,* where his "transcendental argument" appears, Sosa does not dwell on specific details concerning the different ways of understanding naturalism in philosophy; instead, the problem that will concern him is presented in a rather concise way:

> Can the naturalist view us coherently as animals with sensory receptors that enable perceptual and other knowledge of our surroundings? *The brute, blind etiology of our faculties* is said to pose the following problem: "From our own rational point of view, how can we know that we are reliably attuned to our surroundings through our sensory receptors? How then can we properly rely on the deliverances of our sensory mechanisms?" (Sosa, 2010, p. 152 Italics are ours).

> Naturalists must confront *how accidental the success of our cognitive faculties appears to be*, if we go by evolution and by the naturalist conception of our minds as our contentful brains (Sosa, 2010, p. 153. Italics in the original).

So, roughly, here we find Sosa presenting a description of the way our cognitive faculties relate to the world, such that their origin would be "brute" and "blind" and their success in depicting the world would be "accidental". Before turning to Sosa's own "transcendental" solution to the problems raised by this description, we need to take a closer look to the presuppositions implicit here. To begin with, it is not clear why natural selection should be deemed as "brute" and "blind". .*If* we conceded to Sosa that natural selection is "blind" in some relevant aspect, then his starting point could be an argument that, from a premise such that "Our cognitive capacities have evolved in a 'blind' way (that is, in a way that does not warrant that those capacities will be sensitive to reality as it is)" arrives at the conclusion "It would be an accident (a question of sheer luck—and we do not know whether or not we have been lucky) that our cognitive faculties have nonetheless turned out to be sensitive to reality as it is". But this would not be a minor concession; in fact, it would be a rather hasty starting point: the idea that our origins are "blind" and "brute" does not seem to suit the fact that it is the *emergence* of new biological traits which is random, but not their *selection* under the pressures of the environment (cf. Dawkins, 1996, Chapter 3 for a succinct defense of this tenet). Why then should we make this concession about "accidentality" at all? A quick answer could be the following: Sosa does not intend to commit himself with the tenet that our cognitive capacities in fact evolved by accident and that, as a consequence of that, they might be unreliable. On the contrary, the tenet that they

evolved by accident can be seen as a concession *that he himself* makes to a hypothetical interlocutor, in order to prove that *even if* such faculties evolved in that way, we could nonetheless avail of an argument of principle whose conclusion is that such faculties are reliable anyway.

However, even if Sosa makes such a concession, it seems that *something more* should be said so that his interlocutor's doubt can even arise, given what Sosa himself wrote, in a previous text, about similar doubts by Alvin Plantinga. In fact, it was Plantinga who presented the idea that, given a "blind" natural selection, the probability of our cognitive faculties being reliable is either low or unknowable (Plantinga, 1993, p. 231), in an argument that Sosa questioned years before *Knowing Full Well*. In a nutshell, Sosa's point against Plantinga had been that even though it may indeed be rather unlikely that a certain event takes place by sheer chance, this unlikelihood is only relevant if we do not already *know* that the event did take place; the unlikelihood ceases to be relevant if, in fact, we start from the discovery of this surprising event and only need to acknowledge that the unlikely has taken place. If we happened to find a round stone perfectly able to roll downhill, argues Sosa, we might be astounded that such a stone had been "created" by the sheer accident of natural forces; however, he continues, "[w]e can plainly see […] the stone's smooth roundness, and we know through much experience that such an object would roll reliably. The improbability of its having been rounded so smoothly, given its origin in brute forces, is then, surely, no bar to our still knowing it to be smoothly round". And, if the very point of this argument-from-unlikelihood consists on assuming that the emergence of our cognitive faculties is a contingent natural event as any other, then that assimilation turns against the argument itself: "If that is a plausible response", assumes Sosa, "in the case of the round stone, why are we

deprived of it when it comes to our own nature as reliable perceivers endowed with eyes and ears, etc., by means of which we have reliable access to the colors and shapes around us?" (Sosa, 2002, p. 100).

Notably, when in *Knowing Full Well* Sosa presents something similar to an argument-from-unlikelihood, he does not refer to his own dissolution of the problem as it had appeared in Plantinga's work. The point, however, remains: *why is it* that we are supposed *not* to believe that, concerning ourselves just as in the case of the stone, "the improbable has happened" (Sosa, 2002, p. 100)? The actual starting point of Sosa's problem must surely be (and this is only to be expected, given that the last chapter of *Knowing Full Well* is called "Epistemic circularity") that forming beliefs about our own cognitive faculties, *unlike* forming beliefs about round stones, involves a certain self-referentiality: we are referring to the very faculties whose reliable exercise would be required in order to obtain acceptable evidence that they are reliable. In Sosa's own phrasing of the problem (which he will immediately answer himself), "as soon as the implicit belief in the reliability of the faculty is put in question, one can hardly find support for that belief by appeal to the ostensible deliverances of that faculty. Such appeal would after all require a priori trust in the reliability of the faculty, and we would be in a vicious circle" (Sosa, 2002, p. 100). Nevertheless, as is known in the light of other work by Sosa, finding that a certain justification of our beliefs is epistemically circular is not an insurmountable obstacle for this author (Sosa, 1994, 1997b, 1997a)—and, accordingly, what we will find in *Knowing Full Well* may be read as a new expression of this acceptance of circular arguments. As we will see, according to Sosa, when accepting some tenet about how the origins of our faculties make them in fact reliable, we can remain undisturbed by charges that it is circular to do so. Just as Descartes was able

to argue circularly in order to arrive at his proof of a benevolent God that warranted his very reasoning, Sosa stated in his reply to Plantinga that a "naturalist" can also "develop a view of [her]self and [her] surroundings that shows [her] situation to be epistemically propitious" (Sosa, 2002, p. 102); in the same way, as we will see in a moment, we can reject skeptical doubts by concluding that "our faculties do not have *disabling* origins (e.g., ones that involve powerful and systematic deception)" (Sosa, 2010, p. 157). Therefore, in a more exact formulation, what Sosa's argument will attempt to show is that *even if* we regard the origins of our cognitive faculties as "blind and brute" *and* if it may be circular to argue, from the deliverances of those faculties, that they are reliable, it is epistemically legitimate, nonetheless, to insist that they are in fact reliable.

Let us now turn to Sosa's *ipsissima verba* in presenting this solution.

3. SOSA'S "TRANSCENDENTAL ARGUMENT"

Before offering the argument that is the focus of this article, Sosa mentions in passing an alternative solution which requires nuancing his first description of the status of our cognitive capacities in a naturalistic framework (which presented them as the result of sheer chance). This argument is

> a kind of transcendental argument, according to which we could not possibly have contentful attitudes without a lot of built-in truth. The conditions required for acquiring empirical concepts, for example, entail that our application of such concepts could not be too far off the mark. For it is only through adequate

> sensitivity to the presence or absence of perceptible properties that we acquire corresponding concepts of those properties (Sosa, 2010, p. 154).

This is not, however, a strategy that Sosa is interested in analyzing in *Knowing Full Well* (as he points out, the possibilities of this strategy were previously assessed in Sosa, 2009, Chapter 6). At this point, the author moves to an attempt to illuminate the problem of reliability by means of a mental experiment, which in turn is based on an analogy between the assessment of the cognitive capacities of *the species* and those of *the individual*. Sosa presents the situation as follows:

> Suppose we knew of a pill that would most probably disable anyone who takes it. More specifically, the pill induces a persistent illusion of coherent empirical reality. The belief that one *did* take such a pill clashes with the thought that one is still cognitively reliable nonetheless. This thought is true only if one is so lucky. But how could one rationally believe that one is so lucky, absent special reason for so believing? And how could one gain such a reason without vicious circularity? How could one do so, given how likely one takes it to be that one's cognition is disabled? (Sosa, 2010, p. 153)

The reason why a vindication of one's own cognitive abilities could be circular is clear: a solution for the doubts concerning our own reliability could indeed consist on inductively supporting the conclusion "My cognitive capacities are functioning in a reliable way" on the basis of a series of premises that state that our memory and perception

are making us aware of real entities and events (premises such as "I see this table, which is really there", and so on). However, the acceptability of such premises of the inductive inference is in turn supported by the general proposition that describes our knowing capacities as reliable, and it is *this* proposition which need to be able to defend since it has become doubtful under the hypothesis that he have taken *Disablex*, so this kind of inductive justification appears as unacceptable (Sosa, 2010, pp. 154–155).

However, immediately after stating this unacceptability, Sosa adds a solution that appears incompatible precisely with the restriction presented in the previous paragraph, a solution that he will attempt to justify afterwards. According to Sosa, in effect, a premise which we can use to conclude that we have not taken *Disablex* is simply the one by means of which "we manifest our commitment, at least in our intellectual practice, to the claim that our faculties are indeed reliable"—a move we can make because "we are epistemically within our rights in affirming what we already rightfully commit to in practice" (Sosa, 2010, p. 155). This reference to an implicit "commitment" that can later be explicitly formulated will play a role in Sosa's argument in its final version, but, obviously, what has been said up to this point cannot be enough, given that the insistence on the reliability of our capacities *in spite of the possibility of having taken Disablex* still appears as blatant question-begging. It is at this point that Sosa will finally present his core thesis: the circular vindication of our knowing capacities is acceptable simply because the only alternative *is contradictory*; in other words, when we are to assess the reliability of our cognitive capacities, there is a dilemma between either epistemic circularity or inconsistency. Sosa introduces this idea by means of analyzing what it would be like *to obtain evidence* that our capacities are indeed not reliable—in his example, evidence that we have in fact taken *Disablex*.

"Of course," Sosa conceeds, "there are conceivable scenarios where you acquire considerable evidence that you have taken such a pill" (Sosa, 2010, p. 155)—and, we can assume, it is in light of those scenarios that the straightforward reiteration of the tenet that our capacities are reliable is manifestly insufficient. However, the key problem, Sosa warns us, is that

> Even in these scenarios you could hardly be unequivocally justified in believing what they initially suggest, that you have in fact taken the pill. Nor can they even fully justify you in suspending judgment on that question. For, the claim that you have taken any such pill is a self-defeating claim. Both believing that you have taken it, and even suspending judgment on that question, are epistemically self-defeating. The contrary claim, that you have taken no such pill, follows from what is epistemically obligatory […], namely your commitment to denying the universal *un*reliability of your faculties (Sosa, 2010, pp. 155–156).

Sosa defends his solution by means of a rhetorical question about the consequences of answering that our cognitive faculties *are not*, in the moment of that answer, reliable: "How then can we still coherently trust our faculties in sustaining that very answer?" (Sosa, 2010, p. 156). A similar consequence would follow from answering that we *do not know* whether our faculties are reliable: "Even here, how can we coherently commit to *this* attitude while saying that we can't really tell whether, in so proceeding, we are proceeding cognitively aright?" (Sosa, 2010, pp. 156–157). On this basis, Sosa concludes that only "*the confident affirmative can be fully coherent*" (Sosa, 2010, p. 157).

And, in turn, that answer "that gives us reason to draw its deductive consequences, including (a) that we have never taken any disabling pill, and (b) that our faculties do not have *disabling* origins (e.g., ones that involve powerful and systematic deception)" (Sosa, 2010, p. 157).

4. PRECISING THE MEANING OF SOSA'S TENETS: FOUR ASPECTS

In spite of the confident tone of Sosa's prose, his argument, centered on the alleged self-defeating consequences of doubting the reliability of our own cognitive capacities, is less clear than it should. Four points may benefit from a more detailed analysis: *first*, why exactly the skeptical attitude is incoherent; *second*, what "faculties" Sosa refers to in his argument; *third*, how exactly the second of the two "deductive consequences" mentioned by Sosa resembles the first one, and how it is supposed to function as a reply to skeptical doubts; *fourth*, how the attitude of the subject that reflects on her own situation relates to her actual cognitive state, as possibly knowable from a third-person perspective.

4.1. The problem of incoherence

As to the first of our four points, some work of clarifying it has been done by Ram Neta. In an article discussing a variety of anti-skeptical arguments available for an internalist view of justification, Neta includes the thought experiment we have just reconstructed, and makes some interesting remarks about it. He points out, following Sosa, that

> it is incoherent to hold beliefs of the form: I may have taken *Disablex*, but I don't know

> whether my reasons for believing that I may have taken *Disablex* are trustworthy reasons. But can we affirm merely the first conjunct, without also having to affirm the second? No. Given the consequences of taking *Disablex*, the second conjunct simply follows from the first conjunct (as asserted by me): if it's true that I may have taken *Disablex*, then it follows that I cannot know whether my reasons for believing that I may have taken *Disablex* are trustworthy reasons (Neta, 2016).

In fact, as Neta points out, the kind of inconsistency here is neither semantic not pragmatic: here not only the two conjuncts are not merely logically consistent (as in a Moorean paradoxical belief, of the form "p, but I don't believe that p"), but the first conjunct (I may have taken *Disablex*) even *implies* the second (I don't know whether my reasons for believing it are trustworthy) (Neta, 2016). In other words, what Sosa shows in his thought experiment is not a pragmatic paradox, but a case of *epistemic self-defeat*, which takes place when "either the truth of an argument's conclusion or belief in an argument's conclusion defeats one's justification to believe at least one of that argument's premises" (Silva, 2013, p. 579).

4.2. The "faculties" at stake

The second point in need of clarification, but which does not seriously affect the acceptability of the argument, refers to what the faculties are that Sosa refers to when he writes about "trust[ing] our faculties in sustaining [the skeptical] answer" or to "commit[ting] to [certain] attitude". Is he referring to the faculties *of ratiocination* on which we need to rely when we move from certain propositions which we

know perceptively or mnemonically (propositions such as "I accepted to take part in the experiment" or "I am in a lab right now") to the conclusion "I took *Disablex*"? Or is he referring, on the contrary, to *perception* and *memory*, this is, the very faculties that allow us to obtain those propositions on which we support our inference? Sosa is not explicit about this, but we might assume that a good reason for him not to clarify his point is that in both cases his answer could be roughly the same: if we doubt our capacity to make good inferences, then we will have to doubt the reasonableness of the very inference that we are making in order to conclude that we have taken *Disablex*; if we doubt the reliability of our perception or our memory, then we will have to doubt the very support we appeal to when we obtain the premises to infer that our faculties are not reliable. In either case, our epistemic attitude can be characterized as self-defeating.

If we try to understand why this self-defeat takes place, what seems to come to the fore is the following particularity of the mental experiment Sosa is proposing: insofar as what is at stake is the reliability of our cognitive capacities *as a whole*, the empirical reality with respect to which those capacities would have to appear as non-reliable (the empirical reality to which they would not be sensitive enough) is *the same* empirical reality from which we obtained, by means of those capacities, the evidence according to which they are not reliable (for example, because we *see* the written informed consent that we have signed to participate in the *Disablex* experiment). Or, if what is in question is our capacity to ratiocinate, then the connection that may exist between different propositions, and with respect to which our capacity to ratiocinate is no longer functioning in a reliable way (i.e., we might be making bad inferences) is the same type of connection we need to base ourselves on in order to inferentially arrive at the belief that we have taken *Disablex*. If Sosa can describe as "incoherent" the attitude of

the skeptic, this is because the *universal* doubt that the skeptic is proposing about our cognitive capacities does not allow that we make an exception for *some* specific uses of those very faculties, because of which it is contradictory for the skeptic to consider, simultaneously, that those faculties are not reliable but that *in one particular case*—namely, when they provide us with information about their own unreliability—they are in fact reliable. The skeptic's position should not admit such an exception, and, as a consequence, he is contradicting himself in concluding that our capacities are not reliable. We have, consequently, an all-or-nothing position: *either* we assume that our cognitive capacities are in fact disabled because of having taken *Disablex*, but this is a general conclusion that will need to refer to the data provided by our perception, and then there will be no probative value even for the elements on which we were supposed to reach that conclusion, *or* we assume that we actually *can* use those elements provided by our perception and memory (and which suggest that we have taken *Disablex*), but, insofar as we rely on those elements, we need to presuppose that we have *not* taken *Disablex*, so the conclusion that we have in fact taken the pill cannot be what we assume on the basis of those elements, as we initially believed. The only coherent attitude, the only attitude that does not lead us to contradictorily assume that our cognitive capacities are and are not reliable, is, then, that of assuming (at the level of the individual) that we have not taken *Disablex* and (at the level of the species) that our cognitive capacities do not have a disabling origin.

4.3. The contingent character of the argument's conclusion

So far, the argument still seems to hold. However, there is a third, more troublesome, aspect. As to this aspect, it is relevant to notice what Sosa is *not* doing. Sosa is not saying

that, *whatever the origins of our cognitive faculties*, we can deem them as reliable; he is not saying, in other words, that there can be no genetic arguments to cast doubt on the reliability of those faculties. Doing so, in fact, would undermine the analogy he is attempting to establish between the case of doubting the reliability of the faculties of an individual and that of doubting it when it comes to the species as a whole. The analogy requires that, just as we have to declare, "It is not the case that we took *Disablex*; that is a contingent fact that might have happened but did not", we say something similar about the faculties of our species. As a consequence, when he writes that we are entitled to the belief "Our faculties do not have disabling origins", his point is not "No matter what the origin of our faculties is, it cannot be a 'disabling' one", but that even though our faculties *might have had* a disabling origin (i.e., even though the very idea of a disabling origin, according to Sosa, *makes sense*), it is simply not the case that they have had such an origin.

The fact that, concerning the case of the *individual* reliability, the whole assessment of such a reliability depends on a contingent state of affairs (that of not having taken *Disablex*) is underscored by Sosa himself, by means of a detour around an objection. According to this objection, the need to presuppose the reliability of our faculties (failing which our position would be epistemically inconsistent) applies only "[i]nsofar as we are speaking of our cognitive faculties as a whole". However, the objection continues, "why can't we rely on one faculty (or one set of faculties) to question the reliability of another faculty (or another set of faculties)?" (Sosa, 2010, p. 14, n.). In other words, given two faculties, say, memory and perception, we do not need to presuppose that *both* are reliable; we might only presuppose the reliability of memory and, on that basis, question the way our perception is functioning (because things are not looking now as we *remember* them to be). So far, the core point made

by Sosa remains: in this example, nothing would *prove* that our memory is reliable; we would simply have to presuppose it. But Sosa introduces, at this point, an element of interest for our purposes here:

> Reply: Granted, but we still have a transcendental argument in favor of accepting a contingent conclusion, belief of which might have seemed to lie beyond the reach of a priori support (Sosa, 2010, p. 14, n.).

On the one hand, then, the "transcendental" argument involves "a priori support"; *on the other hand*, however, its conclusion remains a *contingent* proposition—it is contingent whether or not we took *Disablex*, and whether or not our species is cognitively amiss. This is an important aspect to remark, because, even though, according to Sosa's argument, we could never *rationally believe* the negation of this contingent conclusion (i.e., that we *have* taken Disablex, that our species *did* take a wrong evolutionary turn), it might well be the case that such negation *is true*, though the discovery of its truth requires an epistemic perspective which is not that of the first person. And this leads us to the fourth and last aspect in need of clarification.

4.4. Point of view of the agent versus objective cognitive situation

As to this fourth point, here the purpose of Sosa's argument is not in need of clarification because he does not refer to it, but because he seems to say *too much*; in other words, because what he says does not seem to fit his general strategy. Let us introduce this point by means of a detour via a comment by Modesto Gómez Alonso:

> [W]hat Sosa points out is that*, from the point of view of the agent, and with independence of what her objective cognitive situation is*, belief is the only rational option. What is important […] is that, far from renouncing to her intellectual integrity, the agent justifies her confidence by preserving her rational consciousness: justified by reason, her confidence is not blind. […] We have a rational right (and a duty) to believe in our rationality (Gómez Alonso, 2019, pp. 46–47. Italics ours).

Now, under this interpretation, Sosa's argument would refer *only* to what the epistemic agent must believe—i.e., whether or not she must believe that she has taken *Disablex*—, not to her "objective cognitive situation"—i.e., whether or not she has actually taken it.

And, in fact, there would be a clear reason to assume that Sosa dissociates in this manner the perspective of the epistemic agent from her "objective cognitive situation"; namely, the fact that one of the results of the internalism/externalism debate has been to show, in Jennifer Lackey's words, that "justification, when it is understood as the property that is necessary and, when added to true belief, close to sufficient for knowledge, has two general components", one of which is objective and the other one subjective (Lackey, 2008, p. 10 Italics in the original). This is to say, we can ask whether a person is actually in a cognitive situation such that she will probably form true beliefs (and being in this condition is an objective affair she may not know about), and we can also ask whether she is reasonable in forming beliefs as she does (a question that requires us to analyze what evidence is available *for her*). *If* Sosa, as Gómez Alonso claims, were only speaking about the subjective aspect of the problem, the only difference between Sosa's

approach and what Lackey describes in this passage would consist on the fact that, whereas Lackey considers the subject's *rationality* as the "subjective" problem and her *reliability* as the "objective" problem, Sosa, in turn, combines the two questions and asks whether *it would be rational for an agent to believe that she is reliable* when she forms a belief. In other words, we need to distinguish *three* questions, not two: whether the agent is rational in believing that p, whether she is reliable in believing that p, and whether she is rational in believing that she is reliable in believing that p. Either way, the problem remains that even if she is *rational in believing* that she is reliable (i.e., that she has not taken *Disablex*) she might be *wrong* in this rational belief—that is, she might *actually have* taken Disablex. And, according to Gómez Alonso's interpretation, as we have seen, Sosa's argument only refers to what the agent is justified in believing, *not* to her actual cognitive status as a possible victim of *Disablex*. However, things get more complex when we notice that Sosa *does* want his argument not only to refer to the rationality of belief, but also to be "sensitive" to facts concerning the objective cognitive situation of the agent. Let us now turn to that problem.

5. PROBLEMATIZING SOSA'S SOLUTION: THE OBJECTION OF "INSENSITIVITY TO THE FACTS"

As we saw in section 3, Sosa's argument, in its initial statement, is supported by the very strong claim that a conclusion such as "Our knowing capacities are reliable" cannot be defeated by contrary evidence—because any such evidence could only be considered as such on the basis of *relying* on those very faculties. However, Sosa tries to weaken the argument by saying that it does not provide us with "a determinative, *ultima facie* reason". The outcome of this

concession is, as we will see, rather intriguing, because it is not clear how the reasons provided by the transcendental argument can *initially* work against alleged evidence that our capacities are not reliable, and, *subsequently*, be potentially defeated by new evidence that seems to be of the same sort of that evidence the argument initially invited us to exclude. This tension, in fact, reveals a deeper instability of Sosa's position, namely, between a first-person and a third-person perspective in order to determine the reliability of our capacities. And it appears when Sosa considers the following objection:

> It seems that if there were people who have taken the pill, they should accept this argument too. But then the problem remains: If both people who have and people who haven't taken the pill have no choice but to believe that they have not taken the pill, the argument that we have no choice but to believe that we have not taken the pill does not give us any *reason* to believe that we have not taken the pill (Sosa, 2010, p. 157, footnote).

We can refer to this passage as the objection of "insensitivity to the facts". This objection displaces us from the first-person perspective which the initial statement of the argument invited us to consider, and inquires, instead, about how sensitive to the facts, *as perceived from a third-person perspective*, is Sosa's argument, i.e., to what kinds of results the transcendental argument would lead different subjects, some of whom may have taken *Disablex* and some others not. However, when Sosa initially presented the thought experiment of *Disablex*, a few pages before introducing this objection, the question he posed concerned *what an individual should rationally believe* about her cognitive capacities; at that

point, the question *was* clearly one that required a first-person perspective, thus confirming Gómez Alonso's interpretation. The argument *did not*, at that point, concern the (somewhat less interesting) question of what evidences one epistemic agent may have to determine the possible radical unreliability of *another* epistemic agent who might have taken *Disablex*.

And presenting the problem in relation to a first-person perspective, in turn, was coherent with the purpose of the argument being to serve *as an analogy*, for an assessment of our capacities *as a species*. In fact, whereas a human *individual* can have a first-person perspective on her capacities, but another individual can also regard them from a third-person perspective, in what concerns our *species* we do not have, in principle, the choice of seeing our capacities "from outside"[2]. Consequently, in the same manner in which we might reject a question like "How reliable are our capacities—but from an *objective* point of view, not just *for us*?" as simply a bad question, we could also expect Sosa to reject a question such as "How sensitive to the facts (as viewed from a third-person perspective) is the transcendental argument?" as equally misguided.

Nevertheless, Sosa's actual answer is more concessive than we might have expected. It runs as follows:

> But if we have no choice but to so believe [i.e., that we have not taken *Disablex*], in the sense that this is clearly enough our rationally preferable option (at least in the respect that it is more coherent than its alternatives), why then is this not a "reason" for so believing? Can an option be clearly our best rational option even when we have no reason to take it? Isn't

[2] More on this below, section 8.

> the very fact that it *is* our best rational option a fine reason to take it? *Not necessarily a determinative, ultima facie reason, but a fine reason nonetheless* (Sosa, 2010, p. 157, footnote. Italics ours).

Saying that the reason provided by the "transcendental argument" is not "determinative, *ultima facie*" can only mean that it is *defeasible*. Let us try to see how this can work.

6. THE IMPOSSIBILITY TO "COMBINE" A FIRST-PERSON AND A THIRD-PERSON PERSPECTIVE

As a matter of fact, this concessive answer is surprising because it implies acknowledging that the transcendental argument, which functions on the basis of considerations of coherence that have to be made by the person whose reliability is at stake, has to be limited in its scope (it does not offer "a determinative, ultima facie reason") in the light of alleged possible evidence only available from a third-person perspective. But it is not clear how it would be feasible to *combine* these two kinds of considerations, so that *prima facie* reasons are somehow defeated by *secunda facie* reasons. More specifically, *from whose perspective* such a thing could be done.

In fact, *what might Sosa possibly mean* when he writes that, from the perspective of the very subject S whose capacities are in question, there exists a solution which is simultaneously "more coherent than its alternatives" but *not* "determinative, *ultima facie*"? *What* could possibly count, from this perspective (not, of course, from those of other subjects) as *ulterior* evidence to tackle the question of S's reliability? If *I*, individually, have doubts concerning my cognitive capacities as a whole, or if *we*, as a species, consider that our experience of the world might be entirely illusory,

but we are told that the most coherent epistemic attitude is to reject these doubts, what could count (after we have adopted this anti-skeptical stance) as an *ulterior* reason, a reason that could possibly *defeat* the "transcendental" reasons that initially convinced us that our cognitive capacities are in fact reliable? It is not clear how the kind of reason provided by Sosa's argument might *not* be "determinative", and this is because the new elements of judgment that might be added to that reason would only count as evidence if *we presuppose* that our cognitive capacities, on the basis of which we acquire those elements, are reliable. If the objection of "insensitivity to the facts" has to be neutralized, as Sosa seems to believe, by means of pointing out that the coherence considerations the argument appeals to do not need to be "determinative", that should mean that the person that had trusted the "transcendental argument" but who actually *had* taken *Disablex* can be rationally guided from her error. But, how is this supposed to happen? Any *new* data that the person doubting her cognitive capacities might receive (data that might confirm her that she *did* accept to take part in the experiment, that she *did* take a pill, that she *did not* receive a placebo) would have to be rejected by that person, precisely on the very basis that the transcendental argument presents; it is not clear why these reasons would function in a first instance and then simply not.

This point can also be appreciated by noting that, just as the thesis of our own reliability cannot be *defeated* by new empirical information, it cannot be *strengthened* either. Our doubts concerning whether or not we have taken *Disablex* cannot be dissipated by means of the cognitive psychologist entering the lab and declaring "Don't worry: you were part of the control group; we only gave you a placebo" (just as radical doubts concerning the cognitive faculties of the species cannot be answered by the appearance of empirical data showing that our planet's environment makes it

extremely unlikely that non-reliable cognitive mechanisms are reliable with the survival of a species guided by them). The acceptance of this kind of evidence simply presupposes that we have not previously disqualified it under the suspicion that they are precisely one more example of the kind of coherent illusion that we suffer.

The impossibility of having "non-determinative" reasons for our own reliability, reasons whose evidential weight can be defeated or strengthened by new evidence as to our objective cognitive situation, can become even more apparent if we consider, instead of Sosa's thought experiment of *Disablex*, a possibility associated to cognitively disabling conditions existing in the real world. Let us imagine a person, *A.*, that may or may not be suffering from some kind of delusion that makes her radically misunderstand the available evidence, and who disagrees with her therapist about her own diagnosis—she believes that, without a shade of a doubt, her therapist is wrong. The therapist may try to convince *A.* that, *because of her very condition*, she cannot reliably judge her own cognitive capacities, or anything else: when she thinks that her therapist is wrong in treating her as delusional, it is only her paranoia that makes her think so; she should, consequently, be reasonable and accept her therapist's opinion, based on the evidence available to her—and, given that *A.* is not an expert herself, this evidence may include elements rationally acceptable for her, such as "My therapist believes that I suffer from delusion" and "What a psychiatrist believes about their patients is usually right". *A.* may have reasoned in this way in the past and she remembers it, but she thinks she has overwhelming evidence that, in this particular case, her therapist is wrong. The therapist, of course, will insist that the patient is wrong when assessing this evidence.

Now, it is clear that what is required from *A.* cannot be a coherent attitude: *A.* is being asked to appeal *to the evidence*

available to her in order to draw the conclusion *that she cannot reasonably evaluate the evidence available to her*. Following Sosa's "transcendental argument", she will conclude that the only attitude that she can adopt is to assume that she is reliable enough. The objection of the "insensitivity to the facts" would enter the scene at this moment, pointing out that *A*.'s judgment *might actually be* rather unreliable, and that, once again, the transcendental argument that allows her to consider herself reliable is insensitive to the facts—it can *also* be used by, say, paranoid schizophrenics. But, once again, this does not prove that, for *A*. herself, the reasons provided to her by the "transcendental argument" are not "determinative". They *are* in fact *ultima facie* reasons for her—it is only from the point of view of *other* agents that the question of A.'s reliability in determining whether *p* is decided by further considerations (considerations such as "Well, she is a schizophrenic anyway"). We may say, then, that *A* is reasonable in believing that she is reliable, but that A is not (meta-)reliable in believing that she is reliable: she believes that she is reliable because she appeals to Sosa's transcendental argument, and we know that the argument is insensitive to the facts. Nevertheless, her lack of (meta-)reliability may be a fact available *to us*, not to A. herself.

7. A DILEMMA

Now, what is the image that we can form of Sosa's argument once we have considered the tension we have analyzed in sections 5 and 6, between a first-person and a third-person perspective? One possibility could be to

declare[3] that Sosa's argument remains too "idealistic", namely, that it concedes too much to the subjective perspective of the knowing subject whose cognitive faculties are at stake, thus failing to pay enough attention to the fact (on which a "realist" would insist) that the subjective and the objective aspects of the problem do not coincide and that, as a consequence, the subject's belief that her cognitive faculties are reliable could (although rationally justified) easily be *false*. However, Sosa's argument, in its final version, could *also* be objected from the opposite point of view, namely, by saying that the argument pays *too much* attention to the subject's objective cognitive situation, which should simply not be relevant if what the argument is concerned about is not the truth of the conclusion "My cognitive faculties are currently reliable" but the *rationality* of the subject that arrives at this conclusion.

Consequently, the situation could be summarized by stating that Sosa faces, in the end, a dilemma: he should *either* completely give up the transcendental argument, given the objector's remarks, *or* give up the claim that the argument can do justice to such remarks. What Sosa cannot do, in fact, is to have his cake and eat it:

- If (first possibility) the hypothetical objection is to be taken seriously, then it can only mean that the "transcendental" argument is unacceptable, because it provides us with a very poor basis for accepting the conclusion that the subject who goes through the argument currently has reliable knowing capacities. In this case, the transcendental argument does not provide us with any reason, not even "prima facie". The objection states that the basis provided by the "transcendental" argument could

[3] Introducing this first possibility was the valuable suggestion made by an anonymous referee who reviewed the first version of this article. Section 7 is an attempt to do justice to this suggestion.

mislead us into believing something false, i.e., that the considerations of coherence that the argument appeals to are simply not good enough to exclude the possibility that our cognitive faculties are unreliable after all. But in this case, even granting the "transcendental" argument the virtue of providing us with a merely "prima facie" reason would be too much—the argument should be completely rejected.

- If, on the contrary, the "transcendental" argument is considered legitimate, because it proves that even a subject whose capacities were *actually* unreliable could not rationally believe that they were, then the remarks by the hypothetical objector should be seen as misdirected: indeed, this objector would have missed the point of the argument, as one about *rationality* alone, and then there would be no need to weaken the argument as providing merely "prima facie" reasons.

Consequently, a first conclusion of the analysis of Sosa's argument is that, in order to work at all, it *should* be as Gómez Alonso thinks it is (which would fit the situation of our hypothetical subject *A.* considered in section 6); in other words, the argument *should* refer *only* to *the subjective rationality of believing that* one is reliable—which rationality cannot possibly be strengthened or weakened by new information. However, there is a further problem that needs to be developed, which refers to the very possibility of the individual/species analogy that Sosa proposes.

8. THE DISANALOGY BETWEEN THE INDIVIDUAL AND THE SPECIFIC CASE

We have just tried to prove that there is a flaw in Sosa's attempt to weaken his transcendental argument: an individual who is not *actually* reliable cannot avail of reasons to defeat her belief that she is so. However, it is clear, at least, that we *can*, from a third-person perspective, find evidence

that the objective cognitive situation of a certain individual makes her unreliable. It is perfectly possible that we, as human beings, *acquire evidence* that a person (a person which is not ourselves) is *both ("subjectively") rational and ("objectively") wrong* in believing in her own reliability. Concerning the case *of the species*, however, the situation is different: not only our reasons to think that we are cognitively reliable are not defeasible; we cannot in principle *ever* acquire evidence that we were wrong in believing that we were cognitively reliable. Let us try to take a closer look at this problem.

In order to do justice to the way Sosa presents his argument (that is, in order to do justice to Sosa's tenet that the transcendental argument provides us with reasons to consider ourselves reliable, but which do not need to reflect our actual cognitive situation), we need to imagine a situation in which an epistemic agent (analogous to *A.*'s therapist, as we saw in section 6) says something like "Well, the members of the species *Homo sapiens* are of course *rational* in believing that they themselves are reliable, but we know they are not *in fact* reliable. And the transcendental argument they appeal to is insensitive to the facts, so, predictably, they are not (meta-)reliable in calling themselves reliable". But when we try to consider this situation, an obvious *dis*analogy emerges between the individual and the specific case, thus undermining Sosa's approach. Though we might well try to *imagine* seeing our species "from outside", in a third-person perspective (as God, or the Martians, might possibly do), seeing the entire humanity as a psychiatrist could see her delusional patient, it does not seem that we, as human beings, can ever acquire evidence that would justify us in seeing *our own species* in this way. The reason is the following: if we have evidence that a given set of cognitive capacities, such as belief-forming mechanisms in rats−let us call it CCR−is unreliable, this requires that we can compare the beliefs that result from the use of the set CCR with our own, allegedly

more reliable, view of the world: we say that the set CCR led the individuals that possess it to form the beliefs *p, q,* and r_3 which we take ourselves to know to be false, and it led them not to form the beliefs *s, t* and *u*, which we take ourselves to be true and would have been beneficial for the possessors of CCR to have. Consequently, we need to make use of *other* set of cognitive capacities, those of normal human beings−let us call it CCH−, to form our own beliefs about the world, then compare these beliefs with those formed using CCR, and then conclude that CCR is an unreliable set of cognitive faculties. Our set of faculties, CCH, thus functions as a "standard meter" of what it is to believe reliably. And this implies that we cannot have such a third-person perspective on *our own* capacities as human beings.

In fact, this role of normal human capacities as a measuring standard can be exemplified by recourse to the very kind of studies that has sometimes been used to argue that the scientific evidence casts doubt on our own reliability. Actually, what these studies can, at most, prove, is that, *on the basis of what we ourselves take to know about the world*, we can describe certain belief-forming mechanisms (*different from ours*) as non-reliable. Let us consider the famous (and rather barbaric) experiment by John García *et al.*, which shows that rats which are given a certain kind of food, and then radiated enough to become sick, will tend to avoid eating this kind of food again (Garcia, McGowan, & Green, 1972). According to Stephen Stich, this study proves that "most of the beliefs produced by the innate inferential strategy Garcia discovered are false beliefs. So it is just not true that natural selection favors inferential strategies which generally yield true beliefs" (Stich, 1985, p. 124). Now, what becomes apparent in this study and can be summarized as a principle of "better safe than sorry" is, *at most*, that, according to what *we* know, evolution may favor a certain quantity of *false positives*−we know that the food eaten by the rats *was in fact safe*, and find

that the rats form the contrary belief nonetheless, probably because natural selection tends to favor being excessively cautious. This kind of study can only function on the basis of the presupposition that *we*, unlike the rats, actually have an accurate cognitive access to the world, an access on the basis of which we can discover the *un*reliability of the rats' belief-forming mechanisms. Our very possibility of stating that rats have *false* beliefs requires that we rely upon *our own* access to the world as a source of *true* beliefs. We are taking ourselves as the standard meter to assess the "length" of the rat's beliefs. As a consequence, we just cannot cast doubt on the standard itself[4].

This commits us to saying something stronger: whereas, in the case of the *individual* considered in the *Disablex* thought experiment, saying "My cognitive faculties are reliable" amounts to saying "The contingent fact of taking *Disablex* simply did not take place", in the case of the *species*, on the contrary, calling ourselves "reliable" *does not* amount to saying that some contingent event has or has not taken place. Let

[4] The point can be fully clarified if we consider, not studies about belief-forming mechanisms present in other species, but studies about the frequency of, say, bad reasoning among us human beings (which counts, of course, among other ways in which our cognitive performances can be defective). Can *that* kind of evidence get us closer to doubting our faculties, to increasing the possibility of us humans collectively suffering from an enduring "coherent illusion", so that a reasonable analogy with the subject of the experiments with *Disablex* can be proposed? The answer, again, must be negative. Such studies may show that the right ways of reasoning can perhaps not be displayed *spontaneously* and that, as a consequence, reflective thought can be necessary to *correct* such errors—but, again, this does not show (nor would it be possible to show it) that the normatively right ways of reasoning are outside the reach of our species; on the contrary, it is those *right* ways that we are using in order to deem the *other* ones as faulty.

us return to an aspect we cursorily tackled in sub-section 4.3 above. As we mentioned, Sosa's transcendental argument does not only require us to proceed in a way that the skeptic could dismiss as circular reasoning—it also requires, more specifically, that we offer an argument *that favors contingent conclusions*, namely, "We have not taken a disabling pill", and "Our cognitive capacities do not have disabling origins". Now, the problem is that, whereas saying, for instance, "Our cognitive capacities evolved in such-and-such stages, and under such-and-such causal pressures" may indeed be a contingent, empirical proposition, something like "Our cognitive capacities evolved in such a way *that they are reliable*" could only be contingent if, as in the case of the reliability of the capacities *of an individual*, we could avail of some *independent* standard that functions to determine what it is to be a "reliable" capacity—a standard that, being independent, our actual cognitive faculties can fall short of. In the case of the individual who might have taken *Disablex*, it is contingent whether or not her capacities are reliable, because we have a standard of reliability that does not depend on her current situation—the standard of how normal, non-disabled human beings know the world. When it is our whole species that we are talking about, such independent standard is not available.

9. Conclusion

Let us sum up the results of the previous analyses.

1. *Pace* Gómez Alonso, Sosa's transcendental argument is not only focused on the rationality of an epistemic agent assessing her own capacities; on the contrary, Sosa weakens his argument in order to consider the possibility that the conclusion of the argument does not correspond with the agent's actual cognitive situation.

2. Such a weakening of the argument is not, however, particularly fitting for Sosa's purposes: in fact, in order to refer to non-determinative reasons, Sosa needs to compare them to further reasons, concerning the epistemic agent's objective cognitive situation, which are not available from her own perspective.

3. Furthermore, there is another reason why Sosa's transcendental argument fails: in order to provide us with an *analogy* applicable to the case of *our species'* cognitive faculties, Sosa would need to show that we might possibly acquire evidence, from a third-person perspective, that they are in fact not reliable. But this, in turn, presupposes the availability of a standard, independent of the normal functioning of *our* cognitive faculties as a species, of what it is to be "reliable".

REFERENCES

DAWKINS, R. (1996). *The blind watchmaker*. New York ; London: Norton.

GARCIA, J., MCGOWAN, B. K., & GREEN, K. F. (1972). Biological constraints on conditioning. In A. Black & W. F. Prokasy (Eds.), *Classical conditioning* (Vol. 2, pp. 3–27).

GÓMEZ ALONSO, M. M. (2019). Wittgenstein y el impacto de Sobre la certeza en la epistemología contemporánea. In D. Pérez-Chico (Ed.), *Wittgenstein y el escepticismo. Certeza, paradoja, locura*.

LACKEY, J. (2008). *Learning From Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

NETA, R. (2016). How Holy is the Disjunctivist Grail? *Journal of Philosophical Research*, *41*.

PLANTINGA, A. (1993). *Warrant and Proper Function*. Oxford University Press.

SILVA, P. (2013). Epistemically self-defeating arguments and skepticism about intuition. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *164*(3), 579–589.

SOSA, E. (1994). Philosophical Scepticism and Epistemic Circularity. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *6*, 263–307.

_____ (1997a). How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *85*(2/3), 229–249. Retrieved from JSTOR.

_____ (1997b). Reflective Knowledge in the Best Circles. *Journal of Philosophy*, *94*(8), 410.

_____ (2002). Plantinga's evolutionary meditations. In J. K. Beilby (Ed.), *Naturalism defeated* (pp. 91–102).

_____ (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge*. OUP Oxford.

_____ (2010). *Knowing Full Well*. Princeton University Press.

STICH, S. P. (1985). Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality. *Synthese*, *64*(1), 115–135.