Ernest Sosa has made and continues to make major contributions to a wide variety of topics in epistemology. In this paper I discuss some of his core ideas about the nature of knowledge and scepticism. I start with a discussion of and objections against the safety account of knowledge—a view he has championed and further developed over the years. I continue with some questions concerning the role of the concept of an epistemic virtue for our understanding of knowledge. Safety and virtue hang very closely together for Sosa. All this easily leads to some thoughts on epistemic scepticism and on Sosa’s stance on this.

Keywords: Knowledge, safety, epistemic virtue, scepticism, Sosa.

Ernest Sosa has made and continues to make major contributions to a wide variety of topics in epistemology. In this paper I discuss some of his core ideas about the nature of knowledge and scepticism. I start with a discussion of the safety account of knowledge—a view he has championed and further developed over the years. I continue with some questions concerning the role of the concept of an epistemic virtue for our understanding of knowledge. Safety and virtue hang very closely together for Sosa. All this easily leads to some thoughts on epistemic scepticism and on Sosa’s stance on this.

1. Safety

It is 0.05 am on January 1, 2001. Jack just finished his first letter ever to his old friend Jill. Jack knows the time and date and comes to believe that it took him until the 21st Century to finish his first letter to Jill. There seems no reason to doubt that Jack knows this. Does it matter that Jack is confused about when the 21st Century begins? He thinks it began on January 1, 2000. If that would speak against his knowledge claim, then at least for some time at the beginning of the 21st Century many people, perhaps the majority, did not know that they were in the
21st Century. This seems false. We want to grant all those people this kind of knowledge even if they’re confused about a particular year in the past.

However, the calendaric confusion matters with respect to something else. Jack’s belief that he finished his first letter to Jill in the 21st Century could have been easily false. Not much would have had to be different for his belief to be false: He would just have had to finish the letter 6 minutes earlier. He would still have believed that he finished the letter in the 21st Century but that belief would have been false. Jack’s belief is thus not “safe”, as this is called these days. Knowledge, according to some and especially to Sosa, requires a safe true belief. S’s belief that p is safe just in case

(S) S believes that p ⇒ p (with “⇒” for the subjunctive conditional; cf., e.g., Sosa 1999, Williamson 2000).

Safety theorists are, of course, realistic enough to restrict the subjunctive conditional to close possible worlds, and not to include all possible worlds. We do not have to deal with further details of the account or with additional clauses here.

The example above suggests that safety is not necessary for knowledge: Jack’s belief constitutes knowledge but is not safe. This is an interesting result, given that many epistemologists nowadays adhere to the view that knowledge requires safety. But couldn’t the safety theorist say something in reply?

They could point out that Jack’s belief is not safe but also deny that Jack’s belief constitutes knowledge. One reason could be that his true belief was based on a false assumption. However, if we add a no-false-belief condition for knowledge (cf., e.g., Clark 1963), then we’re excluding many clear cases of knowledge. Someone who knows that they are living in the year 2888 by deducing it from the false belief that the millennium started on January 1, 2000 would thus come out as not knowing that they live in the 29th Century. This seems false. It will also not help much to try to argue that Jack’s justified true belief is gettierized. This would seem ad hoc and, at least, in need of support by further arguments.

A more promising route might be to accept that Jack has knowledge and to deny that Jack’s belief is unsafe and rather to affirm it is safe because a world in which he finished the letter 6 minutes earlier is not close enough to actuality to matter. But why is it not close enough? Not simply because that would save the safety account. If there is a fact of the matter which determines whether a possible world is close to the actual world, then our alternative world would rather come out as very close. Or so it seems. If there is no such fact of the matter, then all the worse for the safety account. Calling a possible world “close” to the

---

actual world would simply be arbitrary. Nobody has so far shown that there is a non-arbitrary closeness metric for possible worlds (cf. Lewis, a, b). Whether there is such a metric or not, in both cases the safety theorist cannot get what he wants.

Should one, perhaps, say that Jack’s belief is safe but just not safely safe (cf. Sainsbury 1997; Williamson 2000: 123–130)? Have I mistaken the lack of second-order safety for the lack of first-order safety? Roughly speaking, a belief is safe but not safely safe just in case it could not have been easily been false, given conditions C, but conditions C could easily not have been met. The problem with this view is that there does not seem to be a way to determine such unique conditions C. Depending on which conditions we choose, we get different answers to the question whether a belief was safe. Jack’s belief was safe, given that he was in the year 2001 when he finished the letter but it was not safely safe because he could have easily been in the year 2000. However, Jack’s belief would come out as unsafe, given that he wrote the letter at some point during the winter 2000/2001. Nothing seems to determine one right description or choice of condition C.

If we do not have a plausible closeness metric for possible worlds, then the safety account hangs in the air because judgments about closeness of possible worlds are arbitrary. But even if we have a non-arbitrary ordering of possible worlds according to closeness to the actual world, we do get results in cases like Jack’s which speak against the safety view. One should therefore conclude that safety is not necessary for knowledge.

Sosa has recently proposed to relativize safety to the “basis” of the belief (cf. 2004: 322, fn.3 and 2007: 25–28): “A belief that p is basis-relative safe, then, if and only if it has a basis that it would (likely) have only if true” (2007: 26). However, it is hard to see how this could help the safety theorist in our counter-example above. Why should one say that the basis of Jack’s belief would be different had he finished his letter just a couple of minutes ago? Certainly, to say that the basis would be different because then his belief would be false would trivialize the safety account. As long as no account is given of the criteria for the identity of a basis in general and how this helps against our counter-example (as well as against similar examples), the doubts about the safety account remain.

2. Closure

There are more problems for the safety account. One has to do with a violation of closure:

(C) If S knows that p and also knows that (p → q), then S knows that q.

More conditions would have to be added but this rough, basic form of a closure principle should be sufficient for our purposes here. Take the following version of Kripke’s red barn objection against Nozick’s sensitivity account of knowledge (cf. also Goldman’s dachshund example in
Mary is in fake barn county; only red barns are exempt—they are never fake. Mary is aware that red barns are never fake but she is not aware that she is in fake barn county. She finds herself in front of a red barn and comes to believe and know that there is a red barn in front of her. Her belief is safe. She can infer from it that there is a barn in front of her. However, since her belief that there is a barn in front of her is not safe (it could have easily been wrong in fake barn county), the safety theorist would have to deny that she knows there is a barn in front of her. This, however, is not plausible. Apart from that, (C) would thus be violated (cf. Kvanvig 2004: 209). Sosa (2004: 292–294) concedes that there is a problem and proposes to relativize safety to a basis (see above). Again, it is hard to see how this could help the safety theorist. What is the difference as to the basis? It cannot be the truth of Mary’s belief that there is a red barn nor that her belief constitutes knowledge—because that would trivialize the safety account. As long as no more promising version of the safety account is available we have good reason to remain sceptical.

3. Probabilistic Safety

Perhaps one should express the intuition behind the safety account not in modal but in probabilistic terms (cf. for a similar move for the sensitivity account Roush 2005). Here is the rough idea: Knowledge that p requires a true belief that p which also meets the following condition (with “P (p/ Bp)” referring to the conditional probability of p, given the subject’s belief that p):

\[(\text{S}^*) \quad \text{P (p/ Bp)} > \text{some suitable value m} \]

or, alternatively,

\[(\text{S**}) \quad \text{P (not-p and Bp)} = \text{some low value m} \ll \text{P (p and Bp).} \]

Kvart (2006) has proposed such an account for the case of perceptual and memorial knowledge. The basic, rough idea is that

\[(\text{K}) \quad \text{P (p/ Bp) >> P (p) (cf. Kvart 2006: 7).} \]

And the value of the left hand side has to be high (10). Kvart has much more to say about this but we can leave it at this basic level here.

Unfortunately, there seem to be innsuperable problems with such an account, too. First, it does not work for beliefs in necessary truths. P (p/ Bp) is always maximal (= 1), in such cases and the belief that p has nothing to do with P (p): P (p/ Bp) = P (p). Hence, (S*) or (S**) do not seem to add anything relevant and interesting to the true belief condition for knowledge in cases of necessary truths. Kvart’s condition (K)

\[\text{Pritchard (2005: 167–168) objects that one has to consider a wider range of possible worlds here which would make Mary’s belief that there is a red barn unsafe. However, this move seems ad hoc. One should also add that if Kripke’s example is a problem for sensitivity accounts, then it is also one for safety accounts. It is often assumed that it is only a problem for sensitivity accounts.} \]
would be false in such cases—as he himself recognizes—because \( P(p/Bp) = P(p) \).

But even if we restrict the above probabilistic conditions to empirical knowledge of contingent propositions (as Kvart explicitly does), problems remain. Couldn’t the probability that Bill is in the 21st Century, given that he believes it be high (=1 or close to 1) just because he spends all (or almost all) of his life in that century? If the identity of a human being (like Bill) is determined by the identity of the parents’ sperm and egg and if there was no way Bill’s parents could have saved the sperm and egg for a different century, then the probability that Bill could have lived in a different century equals or approximates 0. Again, it seems that neither (S*) nor (S**) can do any work additional to the true belief condition for knowledge: \( P(p/Bp) = P(p) \). And again, Kvart’s condition (K) turns out to be false. It is not clear how one could further restrict (S*), (S**), (K) or any other principle of that kind in order to deliver the right results. Apart from all that, the probabilistic version of the safety account would have problems with the closure principle, too, as a probabilistic version of the red barn objection shows (I spare the reader the repetition in this case; but see some ideas in Roush 2005: 41–47).

4. Virtues

Perhaps we should then move the focus away from modal or probabilistic safety to virtues? Is knowledge true belief which results from the exercise of an epistemic virtue, as Sosa has been arguing for quite some time? Sosa (2007: 42) proposes the following definition of knowledge: Knowledge is justified true belief which is also apt (that is, the result of the exercise of an epistemic virtue). One could call this “the JTAB account of knowledge” (cf. also Sosa 1988: 174–184; 1991a: 138–145; 1991b; 1991c; 1992: 85–89; 1994: 29–33; 1997: 419–420; 2007: ch. 2).³

A being has a virtue only if there is the possibility of falling short of the virtue or of the exercise of the virtue. This implies that no being which cannot but behave in a certain way can be virtuous (or not virtuous) with respect to that kind of behaviour. A being which simply cannot help but perceive their immediate environment realistically and correctly would not count as epistemically virtuous (nor as epistemically deficient or “vicious”). To the degree that human perception under normal conditions is very much like that we would not have perceptual knowledge. This, however, seems false. This objection is still a relatively “inexpensive” shot against the safety account. But there are more serious worries.

There are cases where it seems rather clear that the subject has knowledge even though no virtue was exercised. This would also make

³ Credit for a performance has two dimensions: the ability and the effort. We may praise someone for their performance because of the great ability that was exercised even if little effort went into it but we may also praise someone for the remarkable effort in the use of more restricted abilities.
the idea doubtful that the exercise of a virtue is necessary for knowledge. Suppose Joe is thinking hard about some difficult mathematical problem. Suddenly he can “see” the solution. He didn’t exercise special mathematical abilities; rather, the solution “just came to him”, a bit like in Kekulé’s case when—according to some accounts—he suddenly “saw” the structure of the benzene molecule in a dream or dream-like state. Joe thus comes to know the conclusion—without the exercise of an epistemic virtue. Sure, very often there are epistemic virtues in the background in such cases, like in Kekulé’s case. But even then, the epistemic virtues were not exercised (though somehow causally effective). In Joe’s case we can even assume that the relevant epistemic abilities or virtues weren’t even present in the background. But there still seems to be knowledge (though probably failing the standards of professional mathematicians’ knowledge) even without (an exercise of) epistemic virtue. Basic perceptual knowledge might be another case. Furthermore, if Joe forgets everything again after five minutes and never ever manages to reconstruct the solution despite repeated serious attempts, we would be even more inclined to deny that he had the epistemic virtue but we wouldn’t deny that he knew the solution even if only for a short while. Knowledge can be short-lived but the corresponding virtue would be a more stable disposition. There are no virtues on one occasion only.4

What about the sufficiency of apt justified true belief for knowledge? Here is a counter-example. Sue is an expert on Rembrandt. Nobody comes close to her ability of telling whether something is a true Rembrandt or rather an imitation produced by a member of his school. Recently, two paintings have been discovered: one an original Rembrandt and the other one an astonishing copy by a pupil. All the other leading experts had been asked and failed to tell which is which. Sue however uses her very special abilities and, after some time, comes up with the correct answer. She has a justified, true and apt belief which qualifies as knowledge. What, however, if we add the following aspect to our example: Unbeknownst to her, some jokester at the museum has used the latest high tech tools to produce reproductions of the same Rembrandt painting and put them next to the real one and the copy by the pupil. Nobody, not even Sue, can distinguish between the original and its high tech reproduction (if unmarked as such). Suppose these reproductions are next to the two paintings Sue has been examining. Easily, she could have been, by accident, presented with a high tech reproduction. It seems that under such circumstances she doesn’t know that the painting in front of her is an authentic Rembrandt. But it also seems plausible to say that her belief is justified, true, and apt: Hasn’t she used her extraordinary epistemic virtues to arrive at a justified

4 In all these cases, Joe was lucky in some respect and his knowledge was “lucky” knowledge in that respect. (The idea that some knowledge can be lucky in a certain sense might seem quite unorthodox; whether it is and in what way, should be investigated further—but not here).
true belief? But then justified, true and apt belief is not sufficient for knowledge.

One might object that even though Sue has some remarkable epistemic abilities she does not have the “right” or “relevant” ones here: abilities which would enable her to tell a Rembrandt from a high tech reproduction. But what, one would want to ask back, counts as the “right” or “relevant” abilities here? Does the subject need those abilities which lead her to the truth in a non-accidental way? If that is the answer, then it comes with a prize: It is not clear anymore whether reference to epistemic virtues does the work it was supposed to do in the first place, namely to explain the non-accidentality of a given true belief. The accidentality problem would remain even given the exercise of some virtue. More would have to be said here if one wanted to defend a virtue theory of knowledge.

Sosa (2007: 96, fn. 1) holds that in fake barn cases (structurally similar to the Rembrandt case above) only reflective knowledge is missing but not animal knowledge (cf. for this distinction: Sosa 1988: 182–184; 1991b: 240; 1994: 29–30; 1997: 422, 427; 2004: 290–292; 2007: ch. 2 Sosa 2009a: ch. 7). Reflective knowledge that p is justified true and apt (second-order) belief that one’s true and justified first-order belief that p is apt.\(^5\) One can apply this kind of response easily to the counter-example above. However, I don’t find this very plausible: It seems very plausible to say that Sue also lacks “animal” knowledge that it is a real Rembrandt, too.

Perhaps one might want to reply that the conditions for the exercise of the epistemic virtue were not the right ones in Sue’s case. Sosa (2007: 33) introduces condition C:

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.

Perhaps the circumstances were not appropriate in Sue’s case. But why should we say that? Is it because she could have easily been wrong? But then the notion of safety rather than the notion of an epistemic virtue is doing the crucial bit of the work here. And we would be back with the problems for the safety view mentioned above. We also should not say that circumstances for the exercise are only appropriate if the subject gains knowledge under those circumstances. Given that we are trying to understand the nature of knowledge, this move would, again, trivialize the virtue account.

\(^5\) Sosa (2007: 32) also characterizes reflective knowledge that p as second-order animal knowledge: animal knowledge that one has animal knowledge that p. So, reflective knowledge that p is based on animal knowledge of the second order. This might seem a bit odd, if only because it is not clear whether second-order attitudes could be “animal” ones at all.
5. Virtuous Scepticism

Do epistemic virtues help against scepticism? Here is an argument to the effect that one traditional form of scepticism only arises for the more virtuous or reflective. Take the template of Cartesian sceptical arguments (with “o” for an ordinary proposition and “s” for a sceptical proposition):

(1) S doesn’t know that not-s
(2) If S doesn’t know that not-s, then S doesn’t know that o
(3) Hence, S doesn’t know that o.
(2) is based on the assumption that
(4) S knows that (o → not-s).

If S knows that o, then S also knows—given closure (C) and (4)—that not-s. In other words, given closure and (4), if follows that

(2) If S doesn’t know that not-s, then S doesn’t know that o

Let us take a closer look at (4) and take a popular example. George knows that if he has hands, then he is not merely (and thus falsely) dreaming that he has hands. Now, knowledge requires understanding the known proposition. Whoever does not know that merely dreaming that p involves the false belief that p does not understand what merely dreaming is and thus does not understand that if they have hands, then they are not merely dreaming that they have hands. Therefore, they do not qualify as knowing (or even believing) that proposition. Hence, if it is true that

(5) George knows that if he has hands then he is not merely dreaming he has hands
then it is also true that

(6) George knows that if he has hands then he does not falsely believe he has hands when he really has no hands.

This is an interesting result. It shows that the Cartesian sceptical argument only works under the assumption that the subject has second-order concepts and second-order beliefs and can form beliefs about their own beliefs. In other words, at least the traditional Cartesian scepticism presupposes reflectivity. A being which is restricted to first-order beliefs—to animal beliefs—is, ironically, not threatened by this kind of Cartesian scepticism. A lack of ability can save one from (some forms of) scepticism. Reflectivity, however, can "destroy" knowledge. Less would be more and more less.

Sosa (2007: ch. 2) argues that dream scepticism only threatens the possibility of reflective knowledge but not the possibility of animal knowledge. There is a weaker and a stronger interpretation of this claim. The stronger claim says that only those who do not reflect upon their epistemic states can retain their animal knowledge. The weaker claim says that in addition those subjects who do reflect on their epistemic state can retain their animal knowledge, too, and only “lose” their reflective knowledge. There is something to be said in favour of
the stronger claim: If one’s claims to reflective knowledge are threatened by dream scepticism, then this seems to give the subject a good reason to become sceptical with respect to her first-order belief, too. Animal knowledge would collapse together with reflective knowledge.

6. Dreams, Beliefs and Scepticism

But is the assumption that we do believe things in our dreams accurate? Sosa has denied this (cf. Sosa 2005, 2007: ch. 1). According to him, to dream is to imagine something and not to hallucinate or falsely believe it. Sosa makes the distinction between what happens in one’s dream and what happens while one dreams and applies this distinction to beliefs and believing. I still believe that there are no dragons even if I am dreaming about dragons. How could I believe in my dream that there are dragons when I really don’t believe that there are dragons? The assumption of an inconsistency seems unconvincing and forced. Hence, we should rather give up the idea that we do have beliefs in our dreams. Sosa extends this argument from dispositional beliefs to occurrent or manifest thoughts. We don’t think or believe things in our dreams, we rather imagine things.

I don’t want to go further into this imagination model of dreaming but rather discuss the way Sosa uses it as an anti-sceptical weapon (cf. Sosa 2005, 2007: ch. 1, cf. also the exchange between Cohen 2009: 124–125 and Sosa 2009b: 142–143). Consider the claim to know that one is seated. Sosa argues that it is rational to go with the assumption that one is sitting and is not merely dreaming it. Only when one is awake can one ask a question (Am I sitting? Do I know that?) and answer it. When one is dreaming one cannot even ask a question (not to mention answer it). Hence, while awake the rational thing to do is to answer questions like “Am I sitting?” in the positive (when apparently sitting).

It is not quite clear in what sense this is rational. If the subject cannot distinguish between being awake and dreaming, then there is no reason accessible to her to “assume” she is awake and having thoughts and beliefs, raising questions and answering them (cf. Ichikawa 2008, 2009).

But perhaps Sosa’s idea here is rather that the subject could reflect in a decision-theoretic way (where the alternative acts are either to trust or not to trust the appearances, and where the outcomes are epistemic ones):

<table>
<thead>
<tr>
<th>Circumstances</th>
<th>Awake</th>
<th>Dreaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acts</td>
<td>Go with appearances</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>Don’t</td>
<td>bad</td>
</tr>
</tbody>
</table>

6 McGinn (2006: ch. 6) argues for an imagination account of dreaming but in contrast to Sosa he also argues that we do have beliefs (or, at least, quasi-beliefs) in dreams (cf. ch. 7, esp. 110, 112)
According to Sosa, however, neither act is available in the case of dreams: neither can one ask or answer questions nor can one suspend judgment, go with or against appearances. How then should we frame the decision matrix?

We need a more general notion of thought, covering both the things we do when awake (genuine asking and answering of questions, suspending judgment, etc.) and when dreaming (“pseudo-asking”, etc.). Let us call these kinds of thoughts “super-thoughts” (“super-asking questions”, “super-answering questions”, “super-suspending judgment”, etc.). Now, one could have problems with such a very general conception of thought—isn’t it an invention of an arbitrary category of thought when we have no reason to assume there is a unitary phenomenon here? On the other hand, if one gives up this idea, then it is even harder to explain what “rationality” could mean in this kind of context; we would, at least, be at a loss when trying to construct a decision-theoretic matrix.

Presumably, the revised version of the matrix for super-thoughts would, according to Sosa, be the following one:

<table>
<thead>
<tr>
<th>Circumstances</th>
<th>Awake</th>
<th>Dreaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acts</td>
<td>Super-go with appearances</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>Don’t</td>
<td>bad</td>
</tr>
</tbody>
</table>

Hence, the first act dominates the second one. In other words, while awake it is rational to go with appearances even if one cannot distinguish between being awake and dreaming.

Still, there are several problems with this strategy. First, why should one assume that there is only one “deviant” or sceptical circumstance? What about evil demons (old and new), brains in vats, Berkeleian worlds, etc.? It seems that we don’t even have an idea of what all the possible deviant circumstances might be. And, upon reflection, we can become aware of that. How then can it be rational then to go with the first act? Second, even if we accept the claim that the circumstances in the matrix above are all there are it is not clear what the outcomes are in the deviant case. Perhaps pseudo-going with appearances will be really bad (because it has an impact on our epistemic virtues when awake)? But then no act dominates the other one anymore. We need an argument which shows that the first act is the rational one in terms of possible outcomes; for that we need an argument which tells us why we should expect the above outcomes rather than others. Third, even if from the perspective of the subject it is rational to super-go with the appearances, it is doubtful whether this is sufficient for a successful reply to the sceptic: One can be justified in one’s false propositional attitudes, even if they are systematically false. The subject could thus be perfectly rational in super-going with the appearances but might still be wrong. This possibility seems sufficient to entertain legitimate sceptical worries.
7. Conclusion

Much more could be said about Sosa’s ideas about knowledge and scepticism and about the objections raised above. But it is better to stop here and see what can be said in Sosa’s defence.\footnote{For discussion and remarks on earlier drafts I am grateful to Ernest Sosa and Joe Morrison.}

References


