

Sosa on Safety and Epistemic Frankfurt Cases

Juan Comesaña

1. Introduction

Much work in epistemology in the aftermath of Gettier's counterexample to the justified true belief account of knowledge was concerned with the possible existence of *modal* conditions on knowledge, conditions which related the proposition believed with the fact that made it true. Sosa was an adherent of such a position, putting forward a *safety* condition on knowledge. Lately, however, Sosa's stance with respect to safety has become much more subtle. This paper has two aims. First, I want to trace the evolution of Sosa's thought on safety. Second, I want to examine the issue of whether there are epistemic analogues of Frankfurt cases. The two aims are connected: I suggest that we should interpret Sosa as having abandoned the safety condition altogether because of the existence of epistemic Frankfurt cases.

2. The Problems With Sensitivity

Robert Nozick (Nozick, 1981) famously proposed the following definition of propositional knowledge:

S knows that p if and only if:

1. p is true.
2. S believes that p.
3. If p were false, S wouldn't believe that p.
4. If p were true, S would believe that p.

Three related features of this definition gained the most attention. First, it entails (given even a moderate non-skepticism) that knowledge is not closed under competent deduction. Second, it

affords a certain concessive answer to the skeptic. And finally, it ties knowledge to what is known in a way that is theoretically satisfying. These three features are related as follows. Suppose that we think that, if all goes well, I can know that I have hands in virtue of undergoing a characteristic experience as of having hands. Nozick's definition can deliver this result, for I truly believe that I have hands (thus satisfying the first two conditions) and I would still believe that I do have hands if things were slightly different (thus satisfying Nozick's reading of the fourth condition) but I wouldn't believe it if I didn't have hands (thus satisfying the crucial third condition). A little reflection will allow me to also know that if I have hands then I am not a *handless* creature who has the same exact evidence that I do. After all, my having hands *entails* that I am not such a creature. But notice that Nozick's definition entails that I *do not* know that I am not a handless creature who has all my actual evidence. If I were such a creature, I would still believe that I am not, and thus, even if I satisfy all the other conditions, I do not satisfy condition 3 of Nozick's definition. Thus, the skeptic is right that I do not know the negations of skeptical hypotheses, but he is wrong in thinking that I therefore do not know much of anything else. Notice that a crucial component of both of these features (the denial of closure and the concessive answer to the skeptic) has to do with condition 3, which has come to be known as the “sensitivity” condition:

Simple sensitivity: If S knows that p, then if p were false S wouldn't believe that p.

Nozick saw that he needed to modify his definition to take into account cases where the belief is insensitive but only because if the proposition were false one would use a different method to arrive at an opinion. Nozick's own case was the following:

Grandma: Grandma sees his grandson and on that basis believes that he is well. However, were he not to be well, the grandson's parents would hide this fact from her and so grandma would still believe that her grandson is well.

Grandma's belief that her grandson is well fails to satisfy simple sensitivity but nevertheless amounts to knowledge. In response, Nozick relativizes sensitivity to methods of belief-formation, proposing a definition that is for our purposes equivalent to this one:

Same-basis sensitivity: If S knows that p based on r, then if p were false S wouldn't believe that p based on r.

Notice that moving from simple sensitivity to same-basis sensitivity leaves intact those features of Nozick's definition of propositional knowledge that we referred to above. In particular, it is still the case that, according to Nozick's definition (supplemented by a very plausible anti-skepticism), knowledge fails to be closed under known logical consequence.

While Nozick himself (and also Dretske, who proposed a similar characterization of knowledge in (Dretske, 1971)) took the failure of closure and the concessive answer to the skeptic to be features of his definition, many other philosophers took them to be bugs. In addition, even if one thinks that the concessive answer to the skeptic is correct and that, therefore, closure does fail, Nozick's definition entails that closure fails even in cases where *nobody* would think it does. Thus, for instance, the definition entails that I may know a conjunction without knowing one of the conjuncts.¹ Or take the following example whose essentials are due to (Vogel, 1990):

Ice cube: I leave a glass of water with ice outside at Noon on a characteristic July day in Tucson. A couple of hours later, I believe (and know) that all the ice in the glass has thawed. Had the ice in the glass not thawed (an unlikely, but not impossible, event), I would have still believed that it did.

In this case, sensitivity is not satisfied, but I still clearly know that the ice has thawed.

In light of these counterexamples most philosophers abandoned their adherence to Nozick's definition, and with it sensitivity as a necessary condition on knowledge (but see (Adams & Clarke, 2005) for a valiant-but, in my opinion, unsuccessful-defense). However, one feature of

¹ Let one of the conjuncts be a proposition with respect to which I satisfy sensitivity (that is to say, such that if it were false I wouldn't believe it) and the other one a proposition with respect to which I don't, and let it be the case that if the conjunction were false then the sensitive conjunct would be false but the insensitive conjunct would be true. In that case, I may well know the conjunction (for I satisfy the sensitivity condition and, if we choose the case correctly, I can also satisfy the other conditions on knowledge posed by Nozick) but I won't know the insensitive conjunct (precisely because it is insensitive).

Nozick's definition that outlived it was the idea that knowledge must be underwritten by some kind of modal connection between belief and what makes the belief true.

3. Sosa on Safety

Ernest Sosa was one of the authors who thought that the problems with sensitivity didn't impugn the more general project of trying to come up with conditions on knowledge which posited a modal connection between belief and fact. Noticing that conditionals of the sort in play in the sensitivity condition don't contrapose, Sosa proposed that we replace sensitivity with its contrapositive (for different formulations of the sensitivity condition, see (Sosa, 1996; Sosa, 1999; Sosa, 2002):

Safety: If S knows that p, then not easily would S believe that p without its being so that p.

On the basis of examples similar to the ones that motivated Nozick to move from a bare sensitivity condition to a same-basis sensitivity condition, Sosa also proposed that we should replace the bare safety condition in favor of a relativized form of safety:

Same basis safety: If S knows that p based on r, then not easily would S believe that p based on r without its being so that p.

Sosa's proposal has garnered both adherents (see for instance (Williamson, 2000) and (Pritchard, 2007)) as well as critics (see for instance (Neta & Rohrbaugh, 2004) and (Comesaña, 2005)). As we will see, Sosa himself eventually sided with the critics. Although I have contributed to that debate, my primary interest here is not to make a further contribution, but to trace the different roles that the notion of safety has played in Sosa's epistemology over the years, and to then frame Sosa's latest thought on the matter in a way that I find illuminating.

In order to track the next evolution of the role of safety in Sosa's epistemology, we need to turn to (Sosa, 2007). In that book, Sosa presents the latest development of his brand of virtue

epistemology. A true belief is *accurate*. A belief may also display a competence on the part of the believer—it may be the result of a stable disposition housed in the subject to discern the truth in a field which includes the proposition in question. When this happens, Sosa says that the belief is *adroit*. A belief may be accurate without being adroit and it can be adroit without being accurate. It could also happen, of course, that a belief is both accurate and adroit. But even then, a belief may not be all that it can be, epistemically speaking. For a belief may be both accurate and adroit without owing its accuracy to its adroitness. For a belief to amount knowledge, Sosa claims, this is precisely what must happen—and when a belief is accurate *because* adroit, Sosa calls it an *apt* belief.

According to Sosa, then, knowledge is apt belief. Moreover, neither safety nor same-basis safety are necessary or sufficient for aptness, and so safety is not required for knowledge. As I said above, different authors have presented counterexamples to the necessity of safety (even same-basis safety) for knowledge. Sosa's own is as follows:

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 adds that the answer to this question is plausibly “Yes”.²

Sosa, therefore, doesn't think that same-basis safety is necessary for knowledge. However, this doesn't mean that Sosa thinks that safety is completely irrelevant to epistemology. On the contrary, safety can enter into the specification of what it takes for a belief to be apt—for remember than a belief is apt just in case its accuracy is due to its adroitness. What does it take for a belief's accuracy to be due to its adroitness? Sosa presents the following principle:

² Although it is a qualified “Yes,” for Sosa wants to add that I don't have “reflective” knowledge, for I don't know that I know that the surface is red. We will not be concerned with this aspect of the jokester case in what follows.

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. (33)

The aptness principle highlights that there is still a place for safety in Sosa's virtue epistemology. To see what that place is we need to unpack the aptness principle a bit. Following some remarks recently made by Sosa in (Sosa, 2009), I will first characterize when a disposition constitutes a competence, and then interpret the aptness principle in terms of that characterization:

Competence: S's stable disposition *D* to believe that *p* under conditions *C* constitutes a *competence* if and only if not easily would *D* issue in a false belief when used under conditions *C*.

Aptness (interpreted): S knows that *p* only if S's true belief that *p* is attributable to a competence of S.

Notice that it follows from the aptness principle so interpreted that S knows that *p* only if S's true belief that *p* arises from a disposition to believe that *p* under conditions where the operation of that disposition would not easily yield a false belief. Sosa takes those conditions to be the *normalcy* conditions for the operations of the corresponding competence. Thus, for instance, that the lighting is not undetectably special is part of the normalcy conditions for the operation of our faculty of vision.

So: just as there are cases that motivate the move from simple safety to same-basis safety as a necessary condition for knowledge, so too (Sosa agrees) there are cases that motivate a move away even from same-basis safety as a necessary condition for knowledge (for instance, his own jokester case). Safety and sensitivity are conditionals, and the move from simple safety to same-basis safety, like the move from simple sensitivity to same-basis sensitivity, consists in weakening the antecedent of those conditionals. Similarly, the move to a same-basis-under-normal-conditions safety condition

(which is in effect what Sosa is appealing to in claiming that knowledge requires the exercise of a competence) is a further weakening of the antecedent:

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.

At first sight, then, Sosa is simply iterating a familiar strategy: on the face of counterexamples to a subjunctive conditional, weaken its antecedent. However, competence safety is different from the other versions of safety in a crucial respect. Simple safety and same-basis safety can be seen as part of the old project of *defining propositional knowledge* in purely non-epistemic terms. But competence safety includes reference to normalcy conditions, and Sosa seems to be prepared to grant that normalcy conditions cannot be specified independently of knowledge itself. Thus, the normalcy conditions for the operation of a disposition are those under which the disposition would yield beliefs that amount to knowledge. If this really is Sosa's position (and we shouldn't underestimate the exegetical difficulties that arise here), then the move from same-basis safety to competence safety marks a qualitative shift in Sosa's attitude towards safety. Before this shift, safety was seen as a plausible candidate for being part of a subvenient base for knowledge that doesn't make an ineliminable appeal to knowledge itself. After the shift, safety is seen as part of an explanation of what exercising a competence requires, an explanation which must itself ultimately be understood in terms of knowledge. Thus, although there is a role for safety to play in Sosa's virtue epistemology, it is a much more modest role than the one it used to play, and a much more modest role than the one it plays for other defenders of safety such as Pritchard.³

4. Epistemic Frankfurt Cases

³ Williamson himself also adopts a non-reductive attitude towards safety.

Having laid out the evolution of the role of safety in Sosa's epistemology, I want now to show how that evolution can be given a fruitful interpretation in light of a connection between the theory of knowledge and the theory of moral responsibility that has gone mostly unnoticed so far. More specifically, I want to draw a parallel between Sosa's latest position on safety and "actual sequence" views of moral responsibility. The overarching theme will be that Sosa's position can perhaps be interpreted as an actual sequence theory of knowledge, and that just as actual sequence theories of moral responsibility arise as a reaction to Frankfurt cases, an actual sequence theory of knowledge arises as a reaction to "epistemic" Frankfurt cases.

Let us start by remembering, however briefly, the Frankfurt cases and their role in the free will literature. One of the traditional views on the free will problem is "incompatibilism": the position according to which determinism is incompatible with the existence of free will and moral responsibility. One argument for incompatibilism goes roughly as follows:

1. If determinism is true, then nobody could have done anything other than what they actually did.
2. For any action Φ and subject S, if S is morally responsible for Φ -ing, then S could have refrained from Φ -ing.

Therefore,

3. If determinism is true nobody is morally responsible for anything.

Harry Frankfurt argued (in (Frankfurt, 1969)) that, whatever else one may think about the argument for incompatibilism, its second premise is false. He did so by presenting cases that falsify that second premise. Those cases have come to be known as "Frankfurt cases." Here is one:

Frankfurt case: Abe is deliberating about whether to raise his arm. After carefully considering the pro and cons, he decides to go ahead and raise his arm and does so on the basis of that decision. Unbeknownst to Abe, a neuroscientist with stealth access to Abe's brain would have made him decide to raise his arm if Abe had shown any signs of not doing so by himself.

Frankfurt's verdict, which has been widely shared in the ensuing discussion, is that Abe is morally responsible for raising his arm despite the existence of the neuroscientist. The neuroscientist, as they say in the literature, is a mere *counterfactual* intervener—he didn't in fact do anything to Abe's brain. Nevertheless, he *would* have intervened and made Abe decide to raise his arm. Therefore, Abe couldn't have refrained from raising his arm. Thus, Abe was responsible for raising his arm despite the fact that he couldn't have refrained from raising his arm. Therefore, premise two in the argument for incompatibilism is false.

What morals can we draw from the existence of Frankfurt cases? Not, of course, that incompatibilism is false. But if we follow Frankfurt we *can* conclude that premise two in the argument for incompatibilism above is false. For obvious reasons, that premise has been called “the principle of alternate possibilities,” or PAP. Can we draw any additional conclusions? Some philosophers have thought that we can. Consider conditions whose satisfaction depends on what is going on in non-actual possible worlds, such as the condition that it could have rained today. Call such conditions “modal conditions.” Some philosophers have thought that a consequence that we can extract from the existence of Frankfurt cases is that there are no (non-trivial) modal necessary conditions on responsibility. The idea is that not only is PAP false, but that its failure follows from the failure of a more general principle. The more general principle which some philosophers think is also refuted by the existence of Frankfurt cases can be stated as follows:

Modal condition: If S is responsible for Φ -ing, then some non-trivial modal condition is satisfied.

Indeed, according to some philosophers, responsibility depends only on what *actually happens*, and so it can *never* depend on modal conditions:⁴

No modal condition: Whether S is responsible for Φ -ing never depends on the satisfaction of some non-trivial modal condition.

⁴ Frankfurt's own “identification view,” for example, denies that modal conditions are ever relevant (see Frankfurt (1971)).

Now, it is important to realize that there is a position which entails that PAP is false but doesn't entail the no modal condition thesis. That position is the following supervenience claim:

Supervenience: Whether S is responsible for Φ -ing supervenes on the causes of S's Φ -ing.

Compare a Frankfurt case with its corresponding normal case where no neuroscientist is involved. In both cases, the causes of the subject's action are the same—remember that the neuroscientist doesn't actually intervene in a Frankfurt case—but only in a normal case could the subject have done otherwise. Therefore, the supervenience claim entails that PAP is false. However, the supervenience claim arguably doesn't entail the no modal condition thesis. Suppose, for instance, that one holds that (at least sometimes) whether certain counterfactuals obtain will determine whether the appropriate causal relations obtain. In that case, the no modal condition thesis would be violated, because whether someone is responsible for Φ -ing will depend on the obtaining of those counterfactuals. But the supervenience claim will still be satisfied, because the counterfactuals affect responsibility only through their affecting whether the appropriate causal connection obtains.⁵

So much for what the correct reaction to the Frankfurt cases is. I now want to explore whether there can be *epistemic* Frankfurt cases—that is to say, cases which show that an epistemic analogue of PAP is false. I am aware of only a handful of epistemologists who have asked this question. I will briefly discuss two of them, Hetherington and Zagzebski, before turning to my own approach. Towards the end of this section I turn to a brief examination of what John Fisher (a main contributor to the literature on free will) has to say about this issue.

In (Hetherington, 2002), Stephen Hetherington considers and rejects the claim that epistemic justification can be grounded in epistemic responsibility, and in the context of evaluating this idea he asks whether there can be epistemic Frankfurt cases. Hetherington's focus is thus justification and not knowledge, but the issues he considers are close enough to our own to

⁵ See Sartorio (forthcoming). As Sartorio notices, the supervenience claim (like PAP itself) is most plausibly understood as a constraint on the satisfaction of the *metaphysical* conditions on freedom—it is a further issue whether the *epistemic* conditions (such as whether the subject had evidence that bad outcomes would result from his actions) also satisfy a similar supervenience claim.

warrant some discussion. Hetherington's argument against that the claim that epistemic responsibility is grounded in epistemic responsibility can be reconstructed as follows:

1. If epistemic justification is grounded in epistemic responsibility, then a subject *S* is justified in believing *p* only if *S* is epistemically responsible in forming (and maintaining) the belief that *p*.
2. *S* is epistemically responsible in forming (and maintaining) the belief that *p* only if (in the same circumstances *S* is in prior to forming that belief) *S* could have done other than believe that *p*.
3. *S* can be epistemically justified in believing that *p* even if *S* couldn't have helped but believe that *p* (in the circumstances *S* is in).

Therefore,

4. Epistemic justification is not grounded in epistemic responsibility.

Premise 1 is simply a consequence of the deontological view that Hetherington is examining. Hetherington justifies premise 3 by imagining a subject so epistemically conscientious that she cannot help but form the beliefs that are justified by her evidence. Surely we don't want to say that it is this epistemic virtue which robs her of justification. But it is premise 2 that I am most interested in, for premise 2 is an epistemic analogue of PAP. Therefore, the question immediately arises: given that Frankfurt cases refute PAP, couldn't there be epistemic Frankfurt cases which refute the epistemic analogue of PAP? Hetherington asks this questions and his surprising answer is "No." Hetherington's formulation of an epistemic Frankfurt case is the following (404):

As an epistemic subject *x* is aware, he has evidence *e*, providing good support for *p* (and not for any contrary *q*), and on the basis of which he forms the belief that *p*. Seemingly, therefore, he forms that belief in a way which makes him epistemically responsible for his having that belief. However, unbeknownst to him, he had no possibility of not forming that belief, given that evidence. For some agent *y* would have made *x* believe that *p* (without *x* being aware of this intervention) if *x*-in these circumstances-had been about to fail to form the belief that *p*.

Hetherington's argument that, in that case, x lacks knowledge is puzzling. It is the following (405):

An epistemic Frankfurt case hypothesizes the following conjunctive state of affairs: (i) x 's being about to fail to form the belief that p , even though (ii) as he is aware, his evidence makes that belief the one he should form. And (i)-plus-(ii) portrays x as *already* failing to be epistemically responsible (...), as regards believing p in relation to e . He is already ignoring or overlooking or not doing justice to the support provided by e for p , because he is already about to fail to form the belief that p and therefore is already responding to e in a way that fails to do justice to what, as he is aware, is e 's justificatory support for p .

As far as I can tell, Hetherington's argument confuses what happens *in the counterfactual scenario* where x is about to fail to form the belief (something which, let us grant, would show epistemic irresponsibility on x 's part) with what *actually happens*. In the actual case, x can be as epistemically meticulous as you want in forming the belief. Indeed, it may be the case that x *would not* fail to form the belief—maybe x is like the epistemically conscientious subject of Hetherington's previous argument, who can't help but form the right belief. Of course, this doesn't mean that it is *logically* (or even *physically*) impossible for x to fail to form the right belief—it just means that x 's psychological characteristics make it so that he wouldn't. In that case, the fact that *if (against character) he were* to show signs of failing to form a belief y would have made him form it anyway in no way shows that x is in fact epistemically irresponsible. Indeed, one wonders why Hetherington accepts (as he seems to do) that the original Frankfurt cases refute PAP if he rejects that epistemic Frankfurt cases refute his premise 2. If the mere fact that it is *possible* for a subject to fail to form the right belief showed that that subject is not epistemically responsible, then the mere fact that a subject can fail to perform an act would show he is not morally responsible. But, of course, those mere possibilities show neither thing.⁶

⁶ Hetherington distinguishes between responsibility-for and responsibility-in. One can be responsible for forming unjustified beliefs, and so this is not the notion of responsibility that is fit to explain epistemic justification. Rather, the only plausible reduction is that epistemic justification is a matter of being epistemically responsible *in* forming a belief. And while Hetherington admits that one may be epistemically responsible for holding a certain belief even if one couldn't have helped but form it, he seems to think that one cannot be epistemically responsible *in* forming a

So Hetherington's argument that there are no epistemic Frankfurt cases doesn't work. But that doesn't settle the question whether there are any epistemic Frankfurt cases. Maybe Hetherington's abstract description of how such a case would go is not enough to elicit a judgment one way or the other. Another philosopher who has talked about epistemic Frankfurt cases is Linda Zagzebski (Zagzebski (2001)). Her epistemic Frankfurt case (as she calls it) is the following:

Suppose that Jones is very good at identifying vintages of Bordeaux. In particular, she has no trouble distinguishing a '94 Chateaux Margaux from very similar wines. Black knows that Jones is going to be tasting different vintages of Margaux without knowing in advance the year of the vintage she is tasting. He has installed a device in her head that can make Jones believe that the next wine she tastes is a '94 Margaux whether it is or not. (...) When Jones tastes the next wine, if she appears about to judge that it is a '94 Margaux, the device does nothing. But if Jones is about to judge that it is anything else, *the device will interfere with her tasting sensations* and will lead her to think that it is a '94 Margaux. Now suppose that she tastes a '94 anyway and believes it is a '94, and Black's device does nothing but monitor what is going on in Jones's nervous system. Jones's tasting faculties and taste memory are working fine and she comes to have a true belief in the normal way. (148, my emphasis.)

Zagzebski claims, and I agree, that Jones knows that she has just tasted a '94 Margaux. There is some ambiguity, however, on what exactly would happen in a counterfactual scenario where "Jones is about to judge anything else." Zagzebski says that Black's device "will interfere with her tasting sensations," which suggests that if Jones is about to judge that it is another vintage *based on her accurate tasting sensations* indicating that it is another vintage, then the device will interfere and give Jones tasting sensations characteristic of a '94 Margaux (which Jones can recognize). However, Zagzebski goes on to say that there is a difference between her epistemic Frankfurt case and evil demon scenarios. "The F[rankfurt] cases," Zagzebski says,

belief if one couldn't have helped but form it. But taking into account this complication in Hetherington's reasoning does nothing to alleviate my puzzlement—why would the fact that it is *possible* for me to fail to form the right belief show that I am not responsible in forming it?

involve manipulation of agency in a way that does not appear in the skeptical scenarios. The evil demon gives the agent misleading sensory inputs that inevitably result in false beliefs, but the agent's control over her reasoning process is not altered. This is like deception, where the Frankfurt cases are cases of coercion. (149)

That indicates that Zagzebski is *not* thinking of her epistemic Frankfurt case in the way described above. Rather, by saying that the device “interferes with her tasting sensations” Zagzebski seems to mean that the device will take over the cognitive processing (she calls it “reasoning”) that Jones undergoes to arrive at her judgment on the basis of her sensations. Thus, when Jones has a tasting sensation as of a vintage other than '94, the device leaves those sensations alone but makes Jones judge that it is a '94 nevertheless.

The case in question is Zagzebski's case, so she is free to legislate anything she wants about it. But, of course, the interesting question is whether Frankfurt cases *have* to work in that way, by making the device interfere with the subject's reasoning from her evidence rather than directly with her evidence. This question is particularly important for the morals that we can draw from the existence of Frankfurt cases. Zagzebski claims, for instance, that Jones “fails some well-known counterfactual conditions for knowledge,” and she adds that “she *may* fail the Nozick conditions since she would have had the same belief even if it had been false” (149, emphasis mine). In a footnote we are given the reason why Zagzebski qualifies her claim with a “maybe:” if we are thinking about same-basis sensitivity, then it is clear that Jones doesn't fail *that* condition. And it is, after all, same-basis sensitivity that Nozick advances as a condition on knowledge, simple sensitivity being merely a heuristic device en route to the final condition. Now, this is so, of course, only if we are conceiving of the case as Zagzebski wants us to: with the device interfering with Jones's reasoning ability but not with her evidence. If we understand the case differently, so that if Jones is tasting a vintage other than a '94 then the device gives her tasting sensations indistinguishable from those she would have had she been tasting a '94, then we have a straightforward counterexample even to same-basis sensitivity. We don't, however, yet have a straightforward example to safety, because we are supposing that Jones is indeed a Bordeaux

connoisseur, and so not easily would she believe that it is a '94 without its being so that it is a '94. It is not crystal clear, therefore, which PAP-like principle Zagzebski's case will refute.

A third philosopher who has talked about epistemic Frankfurt cases is John Fisher. Before discussing Fisher's thoughts on the matter, however, I want to present my own.

One useful thing to do in discussing the possibility of epistemic Frankfurt cases is to get clear on what principle we are thinking epistemic Frankfurt cases would refute were they to exist. The obvious epistemic analogue of PAP would be the following:

PAPK1: If S knows that p, then S could have refrained from believing that p.

We should be careful about what the principle *doesn't* say: it doesn't say that if S knows that p then it is logically possible for S to have refrained from believing that p. That is of course true, and nothing (a fortiori no epistemic Frankfurt case) could refute it. Rather, just as with PAP itself, we have to understand PAPK1 as saying that, under the circumstances S finds himself in when he knows that p, he could have refrained from believing that p.

We can maybe take a hint from both Nozick's and Sosa's move from simple subjunctive conditions to same-basis subjunctive conditions, and propose then the following version of an epistemic PAP:

PAPK2: If S knows that p based on e, then S could have refrained from believing that p based on e.

Do we have, then, a plausible epistemic analogue of PAP? Are there Frankfurt cases that refute it? Notice, first, that Hetherington's epistemically conscientious subject—who wouldn't believe what is not justified by her evidence, and thus couldn't have refrained from believing what she actually believes—already throws doubt on PAPK2. But maybe some philosophers would think that there is some sense of "could have" such that the epistemically conscientious subject still could have refrained from believing what she actually believes. Is Zagzebski's case a counterexample to PAPK2? No, because in Zagzebski's case the counterfactual scenario where the device kicks in is

one where the subject's evidence differs from the actual case, and so PPK2 holds trivially of Zagzebski's case.

Let us return, then, to Sosa's reasons for abandoning safety as a condition on knowledge. Notice, first, that Sosa's jokester works exactly in the way in which Zagzebski says Black's device doesn't: if the surface weren't red, the jokester would make sure that it would be white but illuminated with red lights. Thus, the jokester guarantees that the subject will believe that the surface is red not by interfering with his reasoning capabilities, but by being ready to give him misleading evidence if needed. We are now in a position to give a deeper explanation for why I have knowledge in Sosa's jokester case. In that case, I couldn't have refrained from believing that I am facing a red surface. The jokester functions as a counterfactual intervener that guarantees this result. However, the jokester is a *merely* counterfactual intervener, for he does not interfere at all in the actual process by which I come to believe that I am facing a red surface. My belief is apt (accurate because adroit) despite the fact that I couldn't (in the circumstances) have refrained from so believing. This suggests that, just as PAP is shown to be false by the existence of Frankfurt cases, so too PPK2 is shown to be false by cases such as Sosa's jokester.

One could, based on the falsity of PPK2, also conclude that no modal condition is necessary for knowledge:

No epistemic modal condition: Whether S knows that p never depends on the satisfaction of some non-trivial modal condition.

But I would warn against such a conclusion. To begin with, consider the fact that most epistemologists would impose a *basing* condition on knowledge: to know that p in virtue of some evidence e, S must base his belief that p on e. Moreover, most epistemologists would also argue that the basing relation must be understood in causal terms (although no agreement exists as to the exact nature of the relation). Remember that the supervenience claim for responsibility doesn't entail the no modal condition because at least sometimes whether certain counterfactuals obtain will determine whether the required causal relations obtains. Therefore, if the obtaining of some specified causal relation is necessary for knowledge (for instance, by being necessary for the obtaining of the basing condition), and if the obtaining of at least some of those relations has

counterfactual consequences, then the no epistemic modal condition will be false. Moreover, if we follow Sosa in thinking that knowledge requires the exercise of a competence and that competences are defined in dispositional terms, then that gives us another reason for thinking that the obtaining of some modal conditions will be necessary for knowledge. Both of those reasons, however, leave a corresponding epistemic supervenience claim untouched:

Epistemic supervenience: Whether S knows that p supervenes on the actual operation of the faculty that produced S's belief that p.

Now, the fact that epistemic Frankfurt cases do not force us to reject epistemic supervenience doesn't mean that we should accept it, for there may be other reasons to reject it. Depending on the details of the epistemology of faculties, the epistemic supervenience claim might be inconsistent with the original Gettier cases, for instance—see Gettier (1963). Moreover, the claim also seems to be inconsistent with the truth condition on knowledge. I note, however, that Sosa in various places seems to suggest both that his faculty-centered epistemology solves the Gettier problem and that the operation of a faculty under normal conditions guarantees the truth of the proposition believed. Whatever one thinks about this, it is nevertheless important to see that just as the fact that (at least sometimes) a modal condition is necessary for responsibility doesn't impugn the supervenience claim, the fact that (at least sometimes) a modal condition is necessary for knowledge doesn't impugn the epistemic supervenience claim.

Let me now put the evolution of Sosa's thought regarding safety against the background of the Frankfurt cases and their possible application to epistemology. Just as many philosophers put forward PAP as a condition on freedom, so many epistemologists put forward some epistemic modal condition on knowledge. The sensitivity and safety conditions, together with the early causal and nomological accounts of knowledge of Goldman and Armstrong, can be seen as versions of this epistemic modal condition. All of these accounts faced more or less serious counterexamples. The significance of some of the counterexamples to safety (including Sosa's own) resides in the fact that they can be seen as epistemic Frankfurt cases. And just as many philosophers thought that the existence of the original Frankfurt cases not only refutes PAP, but also suggests that what really matters for freedom is the "actual sequence" that results in the action,

maybe we should think that the existence of epistemic Frankfurt cases not only refutes safety and sensitivity as conditions on knowledge, but also suggests that what really matters for knowledge is the actual sequence resulting in the belief. Now, we should be careful in interpreting what we mean when we say that knowledge depends only on the actual sequence resulting in the belief, just as we should be careful in interpreting what we mean when we say that freedom depends only on the actual sequence resulting in the action. For remember that, just as whether the action was *actually* caused by some relevant event might well depend on the satisfaction of some modal condition, so too whether the belief *actually* resulted from the exercise of a competence might well depend on the satisfaction of some modal condition. So we cannot conclude from the existence of epistemic Frankfurt cases that the no epistemic modal condition claim is true.

Much of my discussion in this section has been inspired by remarks made by John Fisher (Fisher, 1987). However, there are still important differences between Fisher's approach and the one taken here. Fisher himself advocates an actual sequence approach to moral responsibility (the very phrase "actual sequence" is Fisher's). However, Fisher doesn't clearly and consistently distinguish between the no modal condition thesis and the supervenience claim—I myself owe the distinction to Carolina Sartorio (forthcoming). Fisher also came up with the idea that there are cases in epistemology which are analogous to the Frankfurt cases. He takes Nozick's grandma case to be one of them, however, whereas I wouldn't—rather, as indicated above, that case merely motivates the move from a simple sensitivity condition to a same-basis sensitivity condition. The most important difference between Fisher and I, however, has to do with the epistemic supervenience claim. Fisher thinks that one example from Nozick refutes the epistemic supervenience claim. I, however (together, I would think, with most epistemologists), take the case to be merely another counterexample to sensitivity as a necessary condition on knowledge. The epistemic supervenience claim might still need to be rejected, but whether it should depends on the evaluation of the details of Sosa's virtue epistemology.

5. Conclusion

In this paper I have traced the evolution of Sosa's thought on safety. I have reached an exegetical and a theoretical conclusion. The exegetical conclusion is that we should interpret Sosa as having abandoned the project of defending safety as an enlightening condition on knowledge. The theoretical conclusion is that Sosa's abandonment of safety can be supported by appealing to the existence of epistemic Frankfurt cases. These cases give us reason to think that an epistemic counterpart of the principle of alternative possibilities is false. The falsity of that principle underlies the problems behind many of the modal conditions actually proposed in the literature—but we should still be careful in distinguishing the falsity of this principle from the stronger (and likely false) claim that no modal conditions are necessary for knowledge.

Rereferences

- Adams, F. and Clarke, M. (2005). Resurrecting the Tracking Theories. *Australasian Journal of Philosophy*, 83(2), 207-221.
- Comesaña, J. (2005). Unsafe Knowledge. *Synthese*, 146, 395-404.
- Dretske, F. (1971). Conclusive Reasons. *Australasian Journal of Philosophy*, 49, 1-22.
- Fisher, J. (1987), Responsiveness and Moral Responsibility, in F. Schoeman (ed.), *Responsibility, Character and the Emotions: New Essays on Moral Psychology*, Cambridge University Press.
- Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibilities. *The Journal of Philosophy*, 829-39.
- Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 5-20.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23, 121-23.
- Hetherington, S. (2002). Epistemic Responsibility: A Dilemma. *The Monist*, 85(3), 398-414.
- Neta, R. and Rohrbaugh, G. (2004). Luminosity and the Safety of Knowledge. *Pacific Philosophical Quarterly*, 85, 396-406.
- Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.
- Pritchard, D. (2007). *Epistemic Luck*. Oxford University Press.
- Sartorio, C. (forthcoming), Actuality and Responsibility, *Mind*.
- Sosa, E. (1996). Postscript to 'Proper Functionalism and Virtue Epistemology,' in *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, J. Kvanvig (ed.), Rowman and Littlefield.

- Sosa, E. (1999). How Must Knowledge Be Modally Related to What Is Known? *Philosophical Topics*, 26(1/2), 373-84.
- Sosa, E. (2002). Tracking, Competence and Knowledge, in *The Oxford Handbook of Epistemology*, P. Moser (ed.), Oxford University Press.
- Sosa, E. (2007). *A Virtue Epistemology*. Oxford University Press.
- Sosa, E. (2009). Replies to commentators on A Virtue Epistemology. *Philosophical Studies*, 137-47.
- Vogel, J. (1990). Are there counterexamples to the closure principle?, in *Doubting: Contemporary Perspectives on Skepticism*, Roth and Ross (eds.), Kluwer.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.
- Zagzebski, L. (2001). Must Knowers be Agents?, in *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, Fairweather and Zagzebski (eds.), Oxford University Press.