

IS KNOWLEDGE SAFE?

Peter Baumann

I. SAFETY

Knowledge, so it seems to many, involves some condition concerning the modal relation between the relevant belief and the corresponding fact. Nozick's almost classical Tracking (or: Sensitivity) Account of Knowledge spells this intuition out in the following way (with " \Rightarrow " for the subjunctive conditional):

S knows that p iff

- (1) p
- (2) S believes that p
- (3) Not-p \Rightarrow S does not believe that p
- (4) p \Rightarrow S believes that p (cf. Nozick 1981, 172ff.; for an earlier account along very similar lines, cf. Dretske 1970 and 1971).

The subjunctive conditionals do not cover all possible worlds but just the close ones. Furthermore, a relativization to methods of belief acquisition is necessary—as Nozick's grandmother-example shows: "A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. Clearly, we must restate our conditions to take explicitly account of the ways and methods

of arriving at belief" (Nozick 1981, 179). (3) and (4) would have to be modified along the following lines:

(3a) Not-p and S uses method M to settle whether p \Rightarrow S does not believe, via M, that p

(4a) p and S uses method M to settle whether p \Rightarrow S believes, via M, that p (cf. Nozick 1981, 179).

One would have to make several further additions, for instance that S does not use any further methods in addition to M in just one of the cases (3a, 4a) but not in the other. It is not necessary to go into these details here.

People have come up with different kinds of objections to the tracking account (cf., e.g., the contributions in Luper-Foy 1987 as well as, e.g., Sosa 1999 and Kvanvig 2004; but see also Salerno forthcoming and the large scale defense of a modified Nozickian account in Roush 2005). They seem so severe to many, that some philosophers have started to propose an alternative modal account of knowledge which is supposed not to run into such problems. Most prominent is the safety account according to which Nozick's condition (3) and (4) should be replaced by

(3*)S believes that $p \Rightarrow p$ (cf., e.g., Sosa 1999, Williamson 2000, Pritchard 2005a, b).¹

Since subjunctive conditionals don't contrapose, (3*) is not equivalent to (3). Again, one has to read the subjunctive as saying something about close possible worlds, not all possible worlds.² Also, the grandmother in Nozick's example would not meet (3*) but still know that her grandson is well when she sees him. Hence, one has to take the methods used (looking at him, asking other people, etc.) into account (cf. Sosa 2002, Comesaña 2005 and Pritchard forthcoming). Hence (3*) should be modified along the following lines:

(3a*)S believes, via method M, that $p \Rightarrow p$.

There are also several whistles and bells to add but those complexities can be safely ignored here.³

In the following, the focus will be on the claim that (3*) (or (3a*) for that matter) is a necessary condition of knowledge. The question is: Is knowledge safe? This paper will first present a counter-example (II), then discuss possible replies to it (III, IV, V) and finally point at the underlying basic problems for all modal accounts of knowledge (VI).

II. A COUNTER-EXAMPLE

Consider the following example (a variation of Nozick's Jesse James example: cf. Nozick 1981, p. 193)⁴:

Gottit and Nogood

The following story is from Milleville, a small town in the Wild West. Two notorious bank robbers have been doing business in the area for some time: Frederick P. Nogood and Wilbur Gottit. Their faces are on "Wanted" posters all over the place. They are rivals and don't like each other at all. When Nogood goes to the bank, he uses a perfectly deceptive Gottit mask; when Gottit goes to the bank, he uses a perfectly deceptive Nogood mask. Nobody but

they themselves knows this. One day, Frank is walking around in the streets of Milleville when he suddenly sees a bank robber leave the bank with a bag full of money on his back, shooting back at the bank. Frank happens to look at him and there is no doubt for him: It is Nogood. But it really is Gottit with his Nogood mask on. However, by sheer coincidence Gottit's Nogood mask slips at that very moment, and Frank notices all this. This is extraordinary because something like that only happens this one time to Gottit and never to Nogood.

Does Frank know that Gottit is the robber, and not Nogood? Yes, that much seems clear and beyond controversy. However, Frank's belief is not safe in the sense of (3*) or (3a*). There are close possible worlds in which Frank believes that Gottit just robbed a bank when it was really Nogood with his Gottit mask. Hence, it is not the case that knowledge is safe.⁵ Safety is not necessary for knowledge.

Two kinds of strategies for replies to this objection are conceivable. First, one could try to argue that Frank in *Gottit and Nogood* has no safe belief but also does not know that it is Gottit he is seeing. Second, one could try to argue that Frank does know that but is also safe in his belief that it is Gottit. One can leave aside the possibility to argue that Frank's belief is safe but does not constitute knowledge; this would not add anything relevant in this context. Both kinds of replies attempt to argue that knowledge and safety do not come apart in the way suggested above. First a couple of remarks on the first strategy (III). The main focus of the paper will be on the second strategy (IV–V). This will lead to deeper problems with modal accounts of knowledge (VI).

III. FIRST REPLY: REASONABLE DOUBTS

(a) *The Reply*. It was just claimed that it is uncontroversial that Frank knows it is Gottit he is seeing. How can one say that? *Gottit and Nogood* is the author's invention and has

not been tested against the intuitions of more than relatively few people. However, there is a slightly different example which goes back to Nozick 1981, 193: the Jesse James example (not to be explained here). There has been a lot of debate on Nozick but nobody, it seems, has expressed any doubt that Nozick's case might not be a case of knowledge. Hence, one can feel on safe ground when saying that it is uncontroversial that Frank in *Gottit and Nogood* knows that it is Gottit. A sufficient number of people would or do not disagree with it.

However, this is just a contingent factual claim. Assume that someone offers the following thought: Might Frank not wonder whether it really is Gottit or someone else who is wearing a Gottit mask under the slipping Nogood mask (call someone like that a "double masker")? Does not the mere possibility of such a doubt suffice to show that Frank does not know it is Gottit?

To be sure, the possibility of the simple thought whether it really is Gottit would not be sufficient as such to threaten the claim that Frank knows it is Gottit. The mere possibility that Jack might wonder whether Cairo really is the capital of Egypt does not speak against the claim that Jack does know that. More generally: S can know that p even if it is possible that S might wonder whether p. Rather, the person needs a reason to doubt: in Frank's case, for instance, that it could be a "double masker". It is the possibility of reasonable doubt which has a chance of threatening relevant knowledge claims, so it seems.

If Frank does in fact seriously wonder whether he is seeing a double masker, then he does not know it is Gottit. However, in *Gottit and Nogood* no such thought enters his mind. That is why the safety theorist's reply focuses on the possibility of reasonable doubt. That there is a possibility of reasonable doubt is supposed to mean the following here: While in the actual world the subject does have a reason to doubt but does not entertain the

doubt, there is a possible world in which the doubt is reasonable and the subject does entertain it. It is not supposed to mean that there is a possible world in which there is a reason to doubt which does not exist in the actual world. Now, how reasonable is Frank's possible doubt whether he is seeing a double masker not identical with Gottit?

(b) *A Counter-Reply*. Consider two cases. Mary looks at an apple, under normal conditions. It seems perfectly obvious that she can know that there is an apple even though in some possible circumstances she might raise the strange question whether it could rather be some weird kind of animal that just looks like an apple (a "wapple"). That doubt would be irrelevant insofar as it is not reasonable; there are no wapples or anything like that in Mary's world and therefore also no reason in the actual world for Mary's doubt. Jack, in contrast, runs into the president, having just read a newspaper story about the president's double. Jack has very good reasons to wonder whether it really is the president or rather his double. Even the possibility of such a doubt seems incompatible with knowledge that it is the president.

Is *Gottit and Nogood* more like Mary's or more like Jack's case? It depends on further assumptions one makes about the case. If double masking was a common practice in Milleville back then, then it would be reasonable for Frank to wonder whether he is looking at a double masker. It would, however, not be reasonable if not even the idea of double masking existed at that time (except, perhaps, in the heads of some few). One can add this aspect to *Gottit and Nogood*. One could even describe the case in such a way that—given the facts plus the laws of nature—it is impossible for double-masking to occur in Milleville (it might, for instance, just be too hot there for successful double masking).

But couldn't there be other doubts in Frank's mind? And does that not show that

one can replace the initial reply by the safety theorist with another one? Sure, but it won't help. For every reply like that one can add a further detail to the initial description of *Gottit and Nogood* which would take care of it.

One might try to rescue the safety theorist's reply by arguing that there is a possible world in which it is reasonable for Frank to have the above mentioned thoughts. This is certainly correct but it misses the point. One can always think of very strange possible worlds in which all kinds of things are live options. There is, presumably, a possible world in which there are animals which look just like apples (wapples). In such a world Mary (or her counterpart) would have a reason to doubt whether she is going to bite into an apple or into a wapple. But this is certainly not a threat against the claim that she knows, in the actual world (which does not contain any wapples), that there is an apple in front of her. What the original reply by the safety theorist tries to exploit is not that there is a possible world such that in that world it would be reasonable to have certain doubts (that there is double masking going on, that there is a wapple, etc.). It is rather that there is a possible world which, like the actual world, contains a reason to doubt but, in contrast to the actual world, has the subject entertaining the doubt. However, it is, as argued, not difficult to add further detail to the initial description of Frank's case so that Frank simply has no reason to doubt that it is Gottit and thus clearly knows that it is Gottit.

(c) *An Unwelcome Implication.* Another thought can be added. Suppose the safety theorist's reply goes through. This would, it seems, lead to a very easy and "cheap" scepticism. If Frank can reasonably have the doubts mentioned why can Mary not have reasonably have wapple doubts? But if that is so, then it would be very hard to see how one could avoid extreme sceptical implications. Frank could not know that it is Gottit, Mary could not know that she is looking at an apple, and

so on for all kinds of ordinary propositions people usually think they know. That in itself might not be the problem; however, the problem is that this kind of scepticism would be too "easy" to be convincing. This also makes the thought behind it suspicious.

So much about the first strategy for a reply to the counter-example *Gottit and Nogood*: the attempt to argue that Frank is not safe but also does not know that it is Gottit he is seeing. The following two sections will take a closer look at the second strategy: the attempt to show that Frank does know but is also safe.

IV. SECOND REPLY: CRITERIA OF CLOSENESS

(a) *The Reply.* How else could the safety theorist defend the view that knowledge is safe? She could argue that there is both knowledge and safety in *Gottit and Nogood*. She might fiddle around with the closeness metric and argue that, say, only those possible worlds in which Gottit does not wear a mask when Frank draws his conclusions are close worlds. But how plausible is that? And how ad hoc?

It is necessary to take a closer look at this kind of reply. Suppose, for instance, that the safety theorist argues that only those worlds in which Frank sees Gottit's mask slip count as close worlds. These worlds would, e.g., involve a world in which everything is like in the actual world except that Gottit has one more hair on his right leg, or another world which differs from the actual world only insofar as there is something slightly different happening at the "other end" of the universe. A world, however, in which Gottit's mask does not slip or in which Frank is facing Nogood with his Gottit mask would not count as close. So, what determines closeness here? What determines closeness of possible worlds with respect to the question whether Frank has a safe belief or knows that it's Got-

tit? The examples just mentioned suggest that some differences between worlds (different numbers of hairs on Frank's right leg, etc.) are irrelevant whereas others are not irrelevant. In what sense could they be irrelevant or not, and with respect to what? A first answer would be that some differences are *epistemically* irrelevant whereas others aren't. A world would be close to the actual world if and only if its differences from the actual world are epistemically irrelevant (enough). In other words, a world is close to the actual world if and only if what one could call the "epistemic situation" of the subject is the same in both worlds. But what is epistemically relevant? What is an "epistemic situation" and what makes an epistemic situation "the same"?

A difference in what the subject believes (that it's Gottit, that it's Nogood, etc.) does certainly matter but that is trivially so: According to the safety account, a close world is one in which the subject holds the same belief as in the actual world. A difference in truth value should, however, not be counted as epistemically relevant; otherwise only worlds in which the subject's belief is true would be close and safety would hold trivially. There are other non-starters. To say, for instance, that a difference in the relevant facts would make a relevant epistemic difference leaves the initial question open: What counts as relevant? In addition, it is not hard to see that this view runs into the same kinds of problems as the view according to which the truth value of the belief has to remain constant. For similar reasons, it won't help to say that only those worlds are close in which the subject knows the same propositions as in the actual world. Take a belief in the actual world which is safe and also constitutes knowledge. Worlds in which that belief is false would not be close according to the proposal because the subject would therefore not know exactly the same things as in the actual world. Again, the safety account would be threatened by triviality. Finally, it also does not help to say

that only worlds with the same (or similar enough) initial conditions are close. What are the initial conditions? The conditions a second before Frank's belief-forming process starts? Or 10 minutes before that? And what goes into initial conditions and what not? It seems there are too many answers here. It won't help to identify initial conditions with truth conditions (for the obvious reasons mentioned above). But if all that is hopeless, then talk about initial conditions seems not very promising. So, what is (in a non-trivial way) epistemically relevant and makes up the subject's epistemic situation?

(b) *First Version: Sameness of Internalist Reasons*. Here is an idea: Sameness of the subject's epistemic situation involves sameness of warrant.⁶ "Warrant" is used in a very broad sense here: as including the reasons or justification the subject might have for her belief as well as the methods of belief acquisition she is using (sameness of method is already required by the safety condition; this condition reappears here). According to this idea a world is close to the actual world only if the warrant is the same in both worlds.

Interpreting "warrant" in the sense of "reasons" (or "justifications"), the question is: What could be the relevant difference in reasons? Frank's reason in the actual world for believing it's Gottit would presumably be that he sees something that looks exactly like Gottit (under normal conditions of vision, etc.), namely Gottit. Should one say that the way things look to the subject are relevant here? Gottit certainly looks different with his mask on than without it. According to this proposal a difference in the appearance of those things that make the subject believe what she believes would be epistemically relevant. In other words, sameness of epistemic situation involves sameness of subjective evidence. The term "subjective evidence" is used here along the lines of the following explanation: Two pieces of subjective evidence do not differ if and only if the subject does

not see any difference between them with respect to their character as evidence. More could be said about this but this hint should be sufficient here. If one understands “reason” or “evidence” in this “subjective” or “internalist” sense, then there is no relevant difference with worlds in which Frank believes it’s Gottit because he sees something that looks exactly like Gottit, namely Nogood with his Gottit mask on. But the idea was to restrict everything to worlds in which Gottit’s mask slips or at least to exclude worlds in which Frank sees Nogood with his Gottit mask on.

But hasn’t an important aspect been neglected here? In the first case, Frank also sees Gottit’s mask slip—so perhaps he has an additional reason to believe he is facing Gottit? This additional reason would be absent in masked-Nogood cases. Why should this difference matter when it comes to closeness of worlds? One might argue that Frank’s reasons are better in the first kind of case. But that is not at all clear: One could argue that if one detects an attempt to deceive one has a reason to remain sceptical (perhaps Gottit is wearing masks over masks?) which one does not have if there is no detected attempt to deceive (but see above). One can—for the sake of the argument—go with the assumption that the first reason is better than the second. This would still not suffice for an explanation of why masked-Nogood worlds are not close enough. In both cases Frank’s reasons are very good; it is just that one is better than the other. In both cases Frank’s reasons are good enough to justify his belief that he is seeing Gottit. The defender of the safety theory would have to show that in masked-Nogood worlds, Frank’s reason to believe he is seeing Gottit are not good enough for these worlds to be close to the actual world. It is, to say the least, not clear at all, what this kind of relatively small difference in the quality of reasons could have to do with the closeness or remoteness of the relevant worlds. As long as no argument to this effect is even in sight,

there are very good reasons to count masked-Nogood worlds as close enough. However, if that is so, then Frank’s belief is not safe and the counter-example succeeds.

(c) *Second Version: Sameness of Externalist Reasons.* The safety theorist could use a different conception of reasons according to which Frank would have different reasons in both kinds of worlds: In one case what looks like Gottit is in fact Gottit whereas in the other case it is in fact Nogood with his Gottit mask on. This difference only matters if one uses a less subjective or more “externalist” notion of reasons and evidence according to which a subject can have different evidence even if she cannot tell that there is a difference. The safety theorist would first have to argue for an externalist account of evidence. More important is another point. This view would invite one to exclude worlds in which the subject’s belief is false as not close enough because the belief is false: Why are masked-Nogood worlds not close enough if not because it’s Nogood instead of Gottit? This, however, would trivialize the safety account, again. Other externalist accounts of evidence do not seem much better off. Take the view that a person’s evidence is that person’s knowledge (cf. Williamson 2000, chap. 9). A masked-Nogood world would then not be close because Frank has different evidence or does not know the same things as in the actual world (he does not know it is Gottit because it is Nogood). It would certainly be too strong to demand that the subject has exactly the same knowledge in close worlds as in the actual world (see above). The safety theorist would have to explain which differences as to knowledge and evidence matter and which don’t. No such explanation has so far even appeared on the horizon. Apart from that: To explain knowledge via safety, safety via closeness and closeness via knowledge is not very illuminating (even though it is not viciously circular; cf. Williamson 2000, 100 here).

(d) *Third Version: Sameness of Methods.* Another way to go for the safety theorist would be to insist on the sameness of methods amongst the worlds considered. This has the advantage of being an explicit condition of safety. However, the safety theorist would have to explain what a relevant difference in methods is. She would have to show that Frank's methods in the actual case are relevantly different from those in the masked-Nogood cases. Similar problems as for the argument from sameness or differences of reasons will arise. If "method" is construed subjectively or in an internalist way, then there is no relevant difference between the methods used in both kinds of cases. If the safety theorist favors an externalist view of methods, then it will be very difficult if not impossible to see a relevant difference between the actual case and masked-Nogood cases that does not boil down to a difference in truth values: In one case Frank uses the method of looking at the other person's face and going by looks, when it really is Gottit, whereas in the other cases Frank uses the method of looking at the other person's face and going by looks, when it really is Nogood with his Gottit-mask on. The long and the short of all this is that it is hard to see what the relevant difference between those two kinds of cases could be. A conclusive proof that there is no such difference is not in sight but as long as nobody offers a better explanation of the difference between cases, one is very much justified in believing (given the above) that there is no such difference. Hence, masked-Nogood cases will have to be counted as close enough. The counterexample against the safety theory (*Gottit and Nogood*) goes through.

V. CLOSE FAKE WORLDS

(a) *An Additional Argument.* Here is another way to see why the safety theorist has to let those troublesome additional worlds into the set of close worlds. Consider the following

variation of the well know fake barn example (cf. Goldman 1976, 772–773):

Fake Gottit

Many people do robberies in the Milleville area. All of them (including Nogood) wear non-slipping perfect Gottit masks, except Gottit who usually wears a Nogood mask, except today. Frank happens to see Gottit without his mask (he forgot to bring it to work today).

In this case one would and should deny that Frank knows that it's Gottit. He could have easily been wrong had he run into someone else with a Gottit mask on. In other words, all those worlds in which Frank sees someone else with a Gottit mask on count as close worlds.

Now, the original *Gottit and Nogood* example differs from *Fake Gottit* in two ways only. First, in *Fake Gottit* many more people than just Gottit and Nogood do robberies whereas in *Gottit and Nogood* it is just Gottit and Nogood; one can neglect this difference because, obviously, nothing depends on it (and one could easily modify *Gottit and Nogood* to eliminate this difference: just add the assumption that many masked people do robberies in the area). What about the remaining difference, namely that in *Gottit and Nogood* Gottit wears a slipping mask whereas in *Fake Gottit* he doesn't wear a mask at all? Could this difference explain why all those worlds—one can call them "mask worlds"—in which it's not Gottit but someone else with their Gottit masks on are close to the actual world in *Fake Gottit* but not close to the actual world in *Gottit and Nogood*? It is very hard to see, to say the least, how this could make such a difference.

(b) *Reasons and Methods, again.* One way to go here would be to say that Frank has a different kind of reason to believe it's Gottit in the two scenarios: in *Gottit and Nogood* his reason would be that it was evident *that it's Gottit after the mask slipped* whereas in *Fake Gottit* the reason would be it was evident

that it's *Gottit* when Frank looked into the other's face. The differences are different: In *Gottit and Nogood* the difference between Frank's reasons in the actual case and in the mask cases would be such that the latter cases would not be close; by contrast, in *Fake Gottit* the difference between Frank's reasons in the actual case and in the mask cases would be such that the latter cases would be close.

A parallel kind of argument has already been discussed in a slightly different context above; hence, the following remarks can be brief. Again, there is no doubt that all these reasons are different. Perhaps the kind of reason Frank has in the actual case in *Gottit and Nogood* is better than the other three kinds of reason (namely the kind of reason Frank has in the actual case in *Fake Gottit* or in the counterfactual cases in *Gottit and Nogood* or *Fake Gottit*). This, however, is not at all obvious and would be in need of support by an argument. Even if it is correct, this still would not show that these differences as to quality are such that mask worlds are close in *Fake Gottit* but not close in *Gottit and Nogood*. All those reasons are good reasons, and they all support the subject's belief that it is *Gottit* he's seeing. The safety theorist is in need of an argument that shows that the differences between the four kinds of reasons (in the two cases) has the implications for closeness and remoteness of worlds which he needs in order to respond to the counter-example *Gottit and Nogood*: namely that mask worlds are close in *Fake Gottit* but not in *Gottit and Nogood*.

It will also not help much to argue that the methods Frank uses in the actual and non-actual cases in *Gottit and Nogood* and in the actual and non-actual cases in *Fake Gottit* are different such that masked-Nogood worlds are not close in *Gottit and Nogood* but are close in *Fake Gottit*. Apart from well-known problems with method individuation, it is not clear why the differences between the methods used should imply that they are relevantly different. But even if one grants

that, it would still be far from clear that such differences would have any implications for questions about remoteness and closeness of worlds. The safety theorist, again, needs to come up with an argument. As long as there is none, one should assume that mask-worlds are either both close or both remote in both *Gottit and Nogood* and *Fake Gottit*. If (as seems plausible) mask worlds are close in the latter case, then they are also close in the former case. But then Frank's belief is not safe and the original counter-example goes through.

VI. DEEPER PROBLEMS

(a) *A Problem*. There is a deeper problem in the background: What determines whether a possible world is close or close enough to the actual world? So far, the arguments of this paper were in line with the idea of a (non-arbitrary) closeness ranking. But what if there is no such ranking? Then things would look even worse, much worse, for the safety theorist—and for obvious reasons: Arbitrary closeness rankings make the safety judgments based on them arbitrary, too. And, indeed, one needs to be very sceptical about the idea of a non-arbitrary closeness ranking. Do only worlds where first class bank robbers bring their masks to work qualify as close worlds (in *Gottit and Nogood*)? Or does one also have to let those worlds in in which they forget their masks? Are only those worlds close where Frank or whoever watches the scene gets a good view of the bank robber? Or does one also have to consider worlds in which lighting conditions are really bad? There does not seem to be an answer to such questions, at least, not a clear and straightforward answer. It is remarkable that safety theorists or, more generally, epistemologists who propose a modal condition for knowledge usually don't even raise this question of what determines closeness of possible worlds. They seem to take it as intuitive and

unproblematic. But it is not (cf. also the hint in Russell 2005, 31–32).

(b) *Context*. Perhaps whether a possible world is close or close enough to the actual world depends on contextual parameters (which ones can be left open here; cf. Lewis 1973, 91–95). This suggests that context determines whether a belief is safe (because it determines what counts as close) and thus also whether it can count as knowledge. This line of thought need not be pursued further here (but cf. Baumann 2005). It does seem to give the defender of (3*) or (3a*) some room to move. She could argue that on her choice of a set of close possible worlds, Frank's belief in *Gottit and Nogood* is indeed safe (see options like the ones mentioned above). This move, however, won't help. Not only does it look like question-begging. It also does not seem to make the assumed indeterminacy of closeness of possible worlds go away. There is room for faultless disagreement ("faultless" in the sense that two persons may disagree about closeness of worlds without one of them being mistaken). However, there is no room for faultless disagreement about the claim that Frank knows (in *Gottit and Nogood*) that Gottit just robbed a bank. This disparity in itself sheds sufficient doubt on (3*) (and on (3a*) as well). If the safety principle were correct, one would not expect cases to be clear with respect to knowledge and unclear with respect to safety. Hence, even if "knowledge" is context-sensitive, too, it is not context-sensitive in the "same way" as closeness and safety: Knowledge and safety can come apart in the sense just explained.

(c) *Being Safely Safe*. But couldn't the safety-theorist argue that Frank's belief (in *Gottit and Nogood*) is clearly safe? It is just that it is not safely safe: It could have easily been the case that his belief were not safe (for instance under conditions more "normal" for Gottit). Isn't there a confusion here of the lack of iterated, higher order safety (cf. Sainsbury 1997, Williamson 2000, 123–130)

with the lack of first order safety in the sense of (3*) or (3a*)? This move is interesting but in the end it won't help much since it does not explain common intuitions in *Fake Gottit*: It seems obvious that Frank's belief is not safe here, not just not safely safe. Why then should Frank's belief in *Gottit and Nogood* only lack second order safety (given the relevant similarities between the two cases)? Apart from that, it does not remove the above mentioned indeterminacy or context-dependency. To be sure, given that Gottit's mask slipped, Frank's belief was safe; it could have easily not slipped but that only means Frank's belief was not safely safe. But what determines what one has to take as "given"? That his masked did slip? That he brought a mask to work? That he brought something to work (a rifle, for instance)? This leads back to the problems mentioned above (cf. for similar examples: Goldman 1986, 45; Peacocke 1986, 142; Neta and Rohrbaugh 2004, 399–400).

(d) *Reliable Indication*. One final attempt to save the safety account deserves to be mentioned shortly. In a different context (having to do with the problem that, apparently, safety is not closed under known entailment), Ernest Sosa has this to say: "What is required for a belief to be safe is not just that it would be held only if true, but rather that it be based on a reliable indication." (Sosa 1999, 149). It is not quite clear whether Sosa wants to add reliable indication as a second condition to the original safety condition (so that safe true belief would not be sufficient for knowledge) or rather replace safety by reliable indication (so that safety would not be necessary for knowledge). In both cases, however, there would be serious problems. Either Sosa's notion of reliability is a modal one: In that case, it is hard to see why and how it can solve problems that the notion of safety cannot overcome. One should also expect problems with the closeness metric similar to the ones discussed above. Or the notion of reliability proposed by Sosa is not

a modal one and can even help solve the problems with the safety account discussed above. But then it would not be clear at all why one should care much about safety and other modal accounts of knowledge.⁷ Apart from that: Similar arguments like the above ones can be constructed for the notion of reliability. It is easy to modify *Gottit and Nogood* in such a way knowledge and reliability come apart in the same way in which knowledge and safety do. Frank's method could so very reliable that it is very hard to explain how he should still fail to know that it is Gottit (first strategy). Or Frank's knowledge could be very unreliable even though he clearly seems to know it is Gottit (second strategy).

The details can be left aside here. Finally, there are well-known problems with reliabilism (like the generality problem; cf. Feldman 1985, Alston 1995); hence, Sosa might just be buying into additional problems rather than solving the original ones (cf. also Comesaña 2005, 397–398 here).

But couldn't one at least use reliable indication as a criterion for closeness of possible worlds? One might think of the following explanation: Worlds are close to the actual world in the relevant ways only if the subject's warrant involves indicators which are reliable to a similar degree. Mask-worlds would not be close in *Gottit and Nogood* because there is a great difference of reliability

(of reasons, methods, etc.) between the actual world and those mask-worlds; in contrast, mask-worlds would be close in *Fake Gottit* because there is no such difference here. This idea won't help much. If "reliability" is taken as a modal notion, like the notion of safety, then the same kind of problems discussed above will come up again (the reader is spared the repetitions here). Apart from that, one would not be able to use a (modal) notion of reliability without having an idea about closeness of worlds. This would, in other words, get the cart before the horse. If, however, the "reliability" is not taken in a modal sense, then the major work for the safety theorist is done by a non-modal notion—in which case it becomes unclear again why one should care much about a modal notion of knowledge in the first place.

VII. CONCLUSION

The overall conclusion is that it is not the case that knowledge is safe: Safety is not a necessary condition of knowledge. This does not mean that the safety-intuition does not capture something important; however, this is also true for Nozick's sensitivity condition. A satisfactory account of necessary conditions of knowledge has to go beyond both. But perhaps it will not even involve a modal condition.⁸

University of Aberdeen

NOTES

1. Since safety requires the belief to be true not just in some close possible worlds but, of course, also in the "closest" possible world, namely the actual world, every safe belief is also a true belief. Hence, one could also, according to a straightforward account, just say that knowledge is safe belief. Whether this view is plausible at all, can be left open here; the interest here is only in safety as an alleged necessary condition of knowledge.
2. Pritchard 2005a, pp. 71, 156, 163, proposes to restrict the subjunctive conditional to most or nearly all close possible world. This variation of the safety condition can be safely neglected in the following.
3. Consider the following case. Jack is unsure whether he exists. He tosses a coin. Heads always triggers a belief that he exists, tails always triggers a belief that he does not exist. The first toss results in

heads and Jack acquires the belief that he exists. This belief is safe but it does not constitute knowledge. Perhaps one has to add the following further condition (cf. Gundersen 2003, 116):

(3a**) S believes, via M, that not-p \Rightarrow not-p.

However, true belief which meets both (3a*) and (3a**) still won't be sufficient for knowledge. Suppose Jack uses dice throwing as a method of settling whether all triangulars are trilaterals: Whenever he throws any one of the six possible numbers he decides that all triangulars are trilaterals. Jack meets both (3a*) and (3a**) (the antecedens of the latter being always false) but his true belief still isn't knowledge because his method is inappropriate. All this, however, can be left aside here because, again, this paper only deals with the idea that safety is necessary for knowledge.

4. Gundersen 2003, 118–119; Blaauw forthcoming, 31–32; Neta and Rohrbaugh 2004, pp. 399–400; Roush 2005, pp. 118–126; and Comesaña 2005, p. 397, offer related examples (cf. also Sosa 2003, p. 159). Gundersen 2003, Blaauw forthcoming, and Neta and Rohrbaugh 2004 just present examples without further analysis or discussion (on Neta and Rohrbaugh see also Luper 2006); they apparently take it for granted that their examples constitute decisive counter-examples against the safety view. Comesaña 2005 discusses objections against his example (cf. pp. 399–401) but only very shortly. Moreover, he does not problematize the idea that there is a determinate closeness-ranking of possible worlds; he rather seems to think that this is no problem at all. Neta and Rohrbaugh 2004, p. 400, agree here. Pryor 2004, pp. 68–71; and Weatherson 2004, p. 377, mention this point but do not go into it at all and assume that one does not have to worry (too much) about it. However, as will become clear soon, this is a huge problem for all modal accounts of knowledge, not just the safety account.—Pryor 2004 is, by the way, arguing that safety is not violated in certain cases of lack of knowledge (cf. pp. 69, 70); in other words, safety is not sufficient for knowledge (cf. also Vogel 2007, p. 83). This paper is not dealing with this point here but rather with one Pryor is not dealing with, namely that safety is not necessary for knowledge.

5. Even if one restricts the safety condition to most close worlds (cf. Pritchard 2005a, pp. 71, 156, 163), this won't help here: There are very many close worlds in which Frank's belief is false.

6. Many would say that warrant is whatever turns true belief into knowledge. However, not everybody agrees that knowledge decomposes into necessary and sufficient conditions and that one can talk in this way about knowledge (cf. Williamson 2000).

7. If one interprets "reliability" in a probabilistic way—which is very plausible—then the non-modal interpretation of "reliability" is much more attractive (cf. Roush 2005), at least as long as one cannot show that the notion probability has a lot to do with the modal notion of possibility (cf., e.g., Gillies 2000 or Hacking 2001).

8. Thanks for comments and discussion go to Sven Bernecker as well as to participants at the 15th Bled Philosophical Conference in 2007 and to anonymous referees.

REFERENCES

- Alston, William P. 1995. "How to Think about Reliabilism," *Philosophical Topics*, vol. 23, pp. 1–29.
- Baumann, Peter. 2005. "Varieties of Contextualism: Standards and Descriptions," *Grazer Philosophische Studien*, vol. 69, pp. 229–245.
- Blaauw, Martijn. Forthcoming. "Contrastivism: Reconciling Skeptical Doubt with Ordinary Knowledge."
- Comesaña, Juan. 2005. "Unsafe Knowledge," *Synthese*, vol. 146, pp. 395–404.
- Dretske, Fred I. 1970. "Epistemic Operators," *The Journal of Philosophy*, vol. 69, pp. 1007–1023.
- _____. 1971. "Conclusive Reasons," *Australasian Journal of Philosophy*, vol. 49, pp. 1–22.

- Feldman, Richard. 1985. "Reliability and Justification," *The Monist*, vol. 68, pp. 159–174.
- Gillies, Donald. 2000. *Philosophical Theories of Probability* (London: Routledge).
- Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge," *The Journal of Philosophy*, vol. 73, pp. 771–791.
- _____. 1986. *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press).
- Gundersen, Lars Bo. 2003. *Dispositional Theories of Knowledge* (Aldershot: Ashgate).
- Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic* (Cambridge: Cambridge University Press).
- Kvanvig, Jonathan. 2004. "Nozickian Epistemology and the Value of Knowledge," *Philosophical Issues*, vol. 14, pp. 201–218.
- Lewis, David. 1973. *Counterfactuals* (Oxford: Blackwell).
- Luper, Steven. 2006. "Restorative Rigging and the Safe Indication Account," *Synthese*, vol. 153, pp. 161–170.
- Luper-Foy, Steven, ed. 1987. *The Possibility of Knowledge: Nozick and his Critics* (Totowa, N.J.: Rowman & Littlefield).
- Neta, Ram, and Guy Rohrbaugh. 2004. "Luminosity and the Safety of Knowledge," *Pacific Philosophical Quarterly*, vol. 85, pp. 396–406.
- Nozick, Robert. 1981. *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press).
- Peacocke, Christopher. 1986. *Thoughts: An Essay on Content* (Oxford: Blackwell).
- Pritchard, Duncan. 2005a. *Epistemic Luck* (Oxford: Clarendon Press).
- _____. 2005b. "Scepticism, Epistemic Luck and Epistemic Angst," *Australasian Journal of Philosophy*, vol. 85, pp. 185–205.
- _____. Forthcoming. "Sensitivity, Safety, and Anti-Luck Epistemology," in *The Oxford Handbook of Scepticism*, ed. John Greco (Oxford: Oxford University Press).
- Pryor, James. 2004. "Comments on Sosa's 'Relevant Alternatives, Contextualism Included,'" *Philosophical Studies*, vol. 119, pp. 67–72.
- Roush, Sherrilyn. 2005. *Tracking Truth: Knowledge, Evidence, and Science* (Oxford: Clarendon Press).
- Russell, Bruce. 2005. "Contextualism on a Pragmatic, not a Skeptical, Footing," *Acta Analytica*, vol. 20, no. 2, pp. 26–37.
- Sainsbury, R. Mark. 1997. "Easy Possibilities," *Philosophy and Phenomenological Research*, vol. 57, pp. 907–919.
- Salerno, Joseph. Forthcoming. "Truth-Tracking and the Problem of Reflective Knowledge," in *Knowledge and Scepticism*, ed. J. Keim-Campbell and M. O'Rourke (Cambridge, Mass.: MIT Press).
- Sosa, Ernest. 1999. "How to Defeat Opposition to Moore," *Philosophical Perspectives*, vol. 13, pp. 141–154.
- _____. 2002. "Tracking, Competence, and Knowledge," in *The Oxford Handbook of Epistemology*, ed. Paul K. Moser (Oxford: Oxford University Press), pp. 264–286.
- _____. 2003. "The Place of Truth in Epistemology," in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. Michael DePaul and Linda Zagzebski (Oxford: Clarendon Press), pp. 155–179.
- Vogel, Jonathan. 2007. "Subjunctivitis," *Philosophical Studies*, vol. 134, pp. 73–88.
- Weatherston, Brian. 2004. "Luminous Margins," *Australasian Journal of Philosophy*, vol. 82, pp. 373–383.
- Williamson, Timothy. 2000. *Knowledge and its Limits* (Oxford: Oxford University Press).