Four worries for safety

Julien Dutant

September 26, 2008

very unpolished draft, do not quote or circulate

Abstract

Several authors have recently defended a safety condition on knowledge according to which one knows only if one's belief couldn't easily have been wrong. The basic understanding of this clause is that S's belief that *p* is safe iff in all close worlds in which S believes that *p*, *p* is true. Here I raise four problems for basic safety: (1) it cannot deal with fake barn cases, (2) it cannot deal with borderline ignorance resulting from vague predicates or concepts, (3) it treats symmetrical cases in an asymmetrical way when necessary truths are involved, (4) it cannot deal with cases of knowledge based on accidentally-gathered evidence. I argue that all worries point out in the same direction: that safety has to be formulated with reference to methods. The consequences are (1) safety is not so different from reliabilism after all; in particular it faces the generality problem just as reliabilism does; (2) safety thus generalized may be sufficient for knowledge.

Several authors have recently proposed a "safety" condition on knowledge, according to which one knows only if one couldn't easily have been wrong. The condition is popular because a) it seems to account for the subject's ignorance in most Gettier cases, b) its formulation is simple, in particular, it does not involve reference to methods or processes (as opposed to reliabilism or Nozick's tracking) and thus seems to dodge the generality problem, c) it preserves closure [Sosa, 1999], d) it makes no concession to skepticism (as opposed to contextualism or Dretkse-Nozick accounts).

In this paper I want to point out four worries for a basic formulation of the safety condition. All four worries seem to me to point in the same direction, namely, that safety has to be reformulated with a reference to methods or processes with which a belief has been formed. I call this new condition *generalized safety*. That reformulation has several consequences: 1) it's less clear how generalized safety differs from a careful formulation of reliabilism and from virtue accounts like Greco's and Sosa's, 2) generalized safety may be sufficient for not knowledge, and not just necessary, 3) the generality problem reappears.

The worries I want to raise for the basic formulation of safety are 1) the condition cannot handle *fake-barn style cases* (contrary to what is often said!), 2) the condition cannot handle *borderline ignorance involving vague predicates*, 3) the condition's *asymmetrical treatement of symmetrical cases* involving necessary truths vs contingent truths, 4) the condition cannot handle cases of evidential luck.

In the first section I will formulate the basic notion of safety. In sections (2)-(5) I'll present the four worries. In section (6)-(7) I formulate and discuss the generalized safety condition which is meant to address them.

1 The basic safety condition for knowledge

Here are several common conditions on knowledge that have gone under the name "safety". (They are typically presented as necessary conditions for knowledge.)

- One could not easily have been wrong about p.
- One would not believe p unless p was true.
- It could not easily have been the case that one believes p and p is false.
- In all close worlds, if S believes p then p.
- In most close worlds, if S believes p then p.
- One is safe from error.

Noticeable differences among the conditions are 1) some conditions are conditions on beliefs (one's belief should not easily be false), while other are conditions on the subject (one should not easily be wrong). 2) most conditions are specifically about one's falsely believing that p; but the first is about being wrong about p (which includes falsely believing that not-p), and the last about being in error in general (which includes falsely believing something else than p). As will become clear later, these differences are significant.

What I will call *basic safety* is the following condition:

- In all close worlds, if S believes that p then p.

It seems to me that this is the most basic formulation of safety, both in terms of it being the first one (Sosa's Sosa [1999], Williamson's), and the most rigorously formulated. At any rate it will be useful to focus on it. For simplicity, we can reformulate it as follows:

Basic safety. $\Box_C(\mathrm{B}p \rightarrow p)$

Where B is the usual belief operator (subscript for the subject omitted), and $\Box_C$ is an alethic necessity operator restricted to close worlds:

$$\text{“}\Box_C p\text{” is true at } w \text{ iff for any world } w' \text{close to } w,\ w' \models p$$

We may or may not assume that the closeness relation between worlds is partly dependent on the conversational context. That does not affect the following worries.

2 First worry: fake barn cases

Smith is driving along an area full of well-imitated fake barns, happens to look at the only real one around, and comes to believe that it is a barn. The usual epistemologists' judgment about such a case is that Smith does not know that that is a barn, because he could easily have been deceived by a fake barn.

But what is exactly the error Smith could easily have made? There are two candidates:

1. He could have happened to look at another building and wrongly believed that that other building was a barn.
2. The thing he is now looking at could have failed to be a barn, and he would then have wrongly believed that it is a barn.

Let a be a nearby fake barn, b be the real barn, and F express the property of being a barn. I am assuming that Smith's has a singular belief about the particular barn is looking at (a). The two possibilities can be formulated as:

1. $\Diamond(\mathrm{B}(Fb) \land \neg Fb)$
2. $\Diamond(\mathrm{B}(Fa) \land \neg Fa)$

But only the second possibility contradicts basic safety. Basic safety implies:

$$\mathrm{K}(Fa) \to (\Box_C(\mathrm{B}(Fa) \to Fa))$$

Thus what we would need in order to show (on the basis of basic safety) that Smith does not know is:

$$\Diamond_C(\mathrm{B}(Fa) \land \neg Fa)$$

Which implies the following:

$$\Diamond_C(\neg Fa)$$

That is, the building a could easily have failed to be a barn. Now, you might think it acceptable that barns are not essentially barns, so that a barn could have been a barn facade. But suppose an analogous case involving fake *sheep*. By sheer luck, Smith is looking at Arnold, the only real sheep in the field. Could Arnold easily have failed to be a sheep? On some essentialist views, Arnold is necessarily a sheep.

The points generalizes. If F is an necessary property of Gs that fake Gs lack (fake Gs being things that are indistinguishable from Gs to the subject in a given situation), then basic safety cannot predict that one fails to know that a is F in a situation where a is a G surrounded by fake Gs, because a could not easily have fail to be F. So unless one assumes essentialism is false, there will be fake-barn style cases with which basic safety cannot deal.[1]

Note that the relevant notion of possibility in safety is alethic. Could we avoid the problem by switching to epistemic possibilities? Even if Arnold is (in fact) necessarily a sheep, we can maintain that it is epistemically possible that he is not a sheep. The question that arises immediately is: whose epistemic possibilities are we considering? If they are the subject's, safety becomes (almost) trivial and circular. If they are the attributor's, we cannot avoid the problem — in the case at hand, we attributors know that Arnold is a sheep, so it's not epistemically possible for us that he isn't. The only epistemic possibilities which could work, as far as I can see, would have to be some impersonal, common-purpose background epistemic space. It would pretty much look like a set of metaphysical possibilities expanded to include impossible worlds.

One way out of the problem is to reformulate the content of the relevant belief as being descriptive. The content of Smith's belief is not *that a is F*, but rather something like *that the building he is looking at is F*. Now even if the building he is looking at is something that is essentially a barn, the satisfier of the description "the building Smith is looking at" could easily have failed to be a barn. (Similarly, if the president of the USA is somebody that is necessarily a man, that does not that the satisfier of the description "the president of the USA" could not have been a woman.) Here as well, that is a radical move: since the fake sheep case generalizes, we would be led to give up the idea that there are singular beliefs.

At bottom, the issue is the following. Basic safety says that a belief about a particular barn is not knowledge if one could easily have had a false belief *about that barn*. But it seems that in the fake barn cases what prevents knowledge is that one could easily have made an error about a similar but distinct object: either another building in the area, or another building that would have been located just where the real barn now is. So what prevents knowledge about p here is a close possible error about p^*, where p^* is a proposition about the similar but distinct object. One way to preserve the formulation of basic safety is thus to build a descriptive-like proposition which is materially equivalent to p in the actual world and materially equivalent to p^* in the relevant possibility. But if one wants to keep the singular proposition, basic safety won't do.

A way to amend safety is to look at beliefs in close propositions, and not just at belief in the same proposition in close situations.

Object-variable safety $K(Fa)$ only if for all a^* sufficiently similar to a, $\Box_C(\mathrm{B}(Fa^*) \rightarrow$

[1] Strictly speaking, one need only assume (in order to make sure that basic safety deals with all such cases) that if two things are indistinguishable in some situation, then they share all their essential properties. That would strip down the list of essential properties to almost nothing but self-identity.

Fa^*).

What will count as "sufficiently similar"? That is going to depend, I take it, on the way in which the subject's belief is currently formed. Roughly, the relevant a^* are going to be objects which are located nearby a in S's perceived space, and presented in the same modality.[2]

3 Second worry: borderline ignorance with vague predicates

The second worry involves vague predicates and borderline ignorance. For ease of formulation I will assume that epistemicism about vagueness is right. Borderline ignorance is the fact that we ignore whether a vague predicate applies to its borderline cases. It seems that a safety condition should account for such ignorance: when applying a vague predicate to a borderline case, one could easily have been wrong. However, basic safety won't do here either.

Suppose the predicate "is a small number" just applies to integers up to and including 9. Let $S9$ be the proposition that 9 is a small number. Basic safety implies that:

$$\mathrm{K}(S9) \rightarrow (\Box_{\mathrm{C}}(\mathrm{B}(S9) \rightarrow S9))$$

Thus in order to deny K($S9$), we need:

$$\Diamond_{\mathrm{C}}(\mathrm{B}(S9) \wedge \neg S9)$$

And thus:

$$\Diamond_{\mathrm{C}} \neg S9$$

That is, that it is possible that 9 is not a small number. But that is not possible: we assumed that being a small number is the property of being 9 or smaller. So basic safety cannot account for the possibility of error in borderline application of vague predicates.

An attractive explanation of ignorance in borderline cases is Williamson's: the phrase "is a small number" could easily have had a different extension without it making any difference to my use of it, in which case I would be just as willing to utter "9 is a small number" even though I would then be saying something false. Let S be the property of being 9 or smaller, and let S^* be the property of being 8 or smaller. The idea is that "being a small number" expresses S but could easily have expressed S^*, and in the latter situation I could

[2]Imagine a being who has a split visual field, one half displaying a location in the US and the other a location in China, both full of similar-looking barn-like buildings. If the creature picks an object at random in her visual field and believes it's a barn, I take it that American fake barns will be relevant to her knowing or not that a Chinese barn is a barn. So perceived space rather than occupied space seems relevant. (Perceived space is not perceptual space: a fake barn just out of the visual field is close in the perceived space, but is out of the perceptual space.)

easily have believed the false content of the sentence "9 is a small number". Thus the relevant possibility of error can be formulated as follows:

$$\Diamond_{\mathsf{C}}(\mathrm{B}(S^*9) \wedge \neg S^*9)$$

which may be the case since $\neg S^*9$ is a necessary truth.

The point holds equally under at least some simple versions of supervaluationism. Suppose that "is a small number" has some range of acceptable precisifications. Let x be the greater definitely small number (*i.e.* the greater number to which all precisifications of "is a small number" apply). Assume x is 9. If you think that we don't know that 9 is a small number, then that cannot be accounted by basic safety, for the same reason as above — 9 could not have failed to be small. If you just think that we don't know that 9 is *definitely* small, then that cannot be accounted by basic safety either — 9 could not have failed to be definitely small either.[3]

The present worry is analogous to the previous one. The previous one stems from the fact that basic safety about Fa requires only to look at situations where one has a belief about a, not at situations where one has a slightly different but (so to speak) subjectively indistinguishable belief about another object a^*. The present worry stems from the fact that basic safety about Fa requires only to look at situations where one has a belief involving the property F, not at situations where one has a slightly different but subjectively indistinguishable belief involving the property F^*.

Here as well, one could be tempted to "go descriptive", this time with the predicates. The content of the subject's belief would be something like that *9 has the property expressed by "is a small number"*. The description "the property expressed by "is a small number"" is satisfied by a different property at different worlds; in some worlds the property picked up is not one that 9 has, so we can recover the idea that the subject's belief could easily have been false.

If one does not want to go that way, the alternative solution is to generalize the safety condition further:

Predicate-variable safety $K(Fa)$only if for all F^* sufficiently similar to F, $\Box_{\mathsf{C}}(\mathrm{B}(F^*a) \rightarrow F^*a)$.

Here as before, whether a predicate is sufficiently similar to another depends on the subject's situation; in particular, in which differences her language could easily have had.

Predicate-variable and object-variable safety can be straightforwardly combined.

[3]One has to be careful here. 9 could easily have satisfied "is not definitely small", since the latter phrase could easily have had a different extension that it actually has. But given the actual extension of the phrase, "9 is definitely small" expresses a necessary truth; in other terms, 9 is necessarily a definitely small number.

4 Third worry: asymmetrical treatment of necessary truths

Consider the two following *lucky guess* cases:

Lucky guess-contingent truth. I write down a rule to select numbers, and I ask Alice whether the number on the board, 97, is one selected by the rule I wrote down. Not bothering even to read the rule, Alice makes a wild guess and answers rightly that it is. (The rule I had written down was: "the selected numbers are the prime ones".)

Lucky guess-necessary truth. I write a number on the board, 97, and ask Bob whether it is prime. Not bothering even to read the number, Bob makes a wild guess and answers rightly that it is.

The datum I assume about both cases is the following: no matter whether they believe what they said, neither Alice nor Bob know the corresponding truth. Alice's truth is *that* 97 *is selected by the rule I wrote down*. It is a contingent truth: I could have had written a different rule, in which case 97 might not have been selected. By contrast, Bob's truth is necessary: it's the truth *that* 97 *is a prime number*.

Let me introduce a second pair of cases:

Bad method contingent truth I write down a rule to select numbers. I ask Celia whether the number on the board, 47, is one selected by the rule I wrote down. Not bothering even to read the rule, Celia adds the digits of the number on the board, and seeing that the result (11) is prime, she rightly answers yes. (The rule I had written down was: "the selected numbers are the prime ones".)

Bad method necessary truth I write a number on the board, 47, and ask Dan whether it is prime. Dan adds the digits of the number on the board, and seeing that the result (11) is prime, he rightly answers yes.

The datum is as before. Here Celia and Dan are using a bad method to answer the question: in both cases, they add up the digits of the number and if the result is prime, they answer "yes" to the question. Celia believes that that is a way to find out the numbers I have selected; Dan believes that that is a way to find out whether a number is prime. In both cases, the method is wrong.

Now, basic safety provides an account of Alice's and Celia's ignorance. In both cases, there is a close situation in which they believe the relevant proposition but the proposition is false: I could have picked up a different rule. But basic safety cannot provide an account of Bob's and Dan's ignorance. In Bob's and Dan's case, the relevant proposition is a necessary truth; and necessary truths trivially satisfy the basic safety condition. Recall the condition:

$\Box_C(Bp \rightarrow p)$

If p is necessarily true, so is $q \rightarrow p$ for any q, and thus the safety condition is satisfied. (The same holds for "easy necessities", truths that hold in all *close* possible worlds.)

Now defenders of basic safety often point out that the safety condition is just meant to be a necessary condition of knowledge, not a sufficient one. Correlatively, they claim that the safety condition is just a condition concerning knowledge of contingent matters. So they are likely not to be moved by Bob's and Dan's case: such case show that we cannot take basic safety to be sufficient, but that was not intended anyway.

But that position is deeply unsatisfactory. It is clear that in Alice's and Bob's cases *the reason for their ignorance is the same*: namely, a lucky guess does not provide knowledge. And similarly in Celia's and Dan's cases: both are applying a method which is a bad one for finding the truth about the present matter. Both pair of cases are symmetrical. And most plausibly, *they call for symmetrical explanations of ignorance.* So we should expect that one condition will both rule out Alice's knowing as well as Bob's, and Celia's as well as Dan's.

The assymetrical treatment that basic safety imposes on such cases suggest that there is a substantial difference between the reasons for Alice's ignorance and the reason for Bob's — but that is *prima facie* implausible needs to be argued for. For instance, suppose we say that Alice fails to know because her belief could easily have been false, while Bob fails to know because he is unjustified. Then it is mysterious why this asymmetry arises; Alice's belief is formed in the same way as Bob is, so isn't it unjustified as well? Then we would have a condition that accounts for both cases of ignorance in a symmetrical way.

If the symmetry considerations are cogent, there is a condition C on knowledge which accounts for ignorance in all cases above.[4]The relation between basic safety and C will be one of the two following:

1. *Basic safety is an under-generalization of C*. That is what the generalized safety condition that we will put forward below attempts to do.

2. *Basic safety is a proper condition on knowledge, independent of C, but it is not the proper explanation of ignorance in the contingent lucky guess / contingent bad method cases.* Here safety would be a proper, independent condition of knowledge. But the proper explanation for ignorance in Alice's and Celia's cases is that subjects fail to satisfy C, not that they fail to satisfy basic safety. The fact that basic safety also predict ignorance in these cases is somehow an accident.

Whichever way turn out to be the right one, defenders of basic safety are left in a bad position: they cannot appeal to the intuitions about lucky guesses and bad methods anymore in order to motivate safety, since such cases are ultimately going to be explained in some other way.

[4]If the account for ignorance differs between bad-methods on the one hand and lucky-guesses on the other, let C be the conjunction of the condition that solve the former and the one that solve the latter.

Let me also point out that the implicit motivation sometimes suggested for the asymmetry is unwarranted. By stating that their analysis is only concerned with "contingent facts", I take it that epistemologists have an implicit picture in mind in which necessary truths are mathematical and logical truths, and contingent truths the empirical ones. The division would thus cut nature at the epistemical joints: *a priori* vs. empirical domains. But that is wrong: assuming some essentialism, a lot of perceptual knowledge is about necessary truths: that Arnold is a sheep, for instance.[5]Vague predicates also generate cases of empirical necessary truths: for instance, pointing at a tree, the fact that *a tree of that size is small*. The case works if we assume that the meaning of "small" picks up absolute sizes and not relative ones; for instance, that it expresses the property of being smaller than 5 meters and not the property of being smaller than most other trees. Given that, suppose that tree T is less than 5 meters: even though it is contingent that T is less than 5 meters — and thus that T is small, it is necessary that a tree of *that* size (the size of T) is small. If you prefer to see "small" as a relative size predicate, you can still build analogous cases with other predicates like "red".

The symmetry considerations show that the fact that basic safety cannot with ignorance of necessary truths is a deep issue which shows that either basic safety is an under-generalization or that the cases that motivate it are ultimately going to be accounted for in some other way. In both case basic safety will prove superfluous.

5 Fourth worry: evidential luck

The fourth worry is well-known but deserves emphasis in the light of the previous ones: basic safety is counter-exampled in cases of evidential luck. Cases of evidential luck are cases in which a subject knows on the basis of evidence she got by luck. Here's one such case:[6]

Granny's hedgehog Granny leaves a cup of food for her cat outside every night and finds it empty in the morning. She happens to wake up one night by sheer accident (a wrong number call) and sees that the animal eating the food is a hedgehog.

Granny now knows that the cat is not eating the food every night. But she could easily have had the false belief that it does: if she hadn't seen the hedgehog, she would still believe that the cat eats the food every night.

There are two ways to deal with such cases:

1. *Amend the closeness relation*. Only the possibilities in which Granny has seen the hedgehog are close.

[5]Note that I am not saying here that we know through perception alone that Arnold is *necessarily* a sheep. I am just saying that we know through perception that Arnold is a sheep; and that happens to be a necessary truth. That is all that is needed for the basic safety condition to be trivialized.

[6]The classic case of that type is Nozick's granny Nozick [1984]. See also Unger's Unger [1968]

2. *Amend the safety condition*. What is required is that *given that she has seen the hedgehog,* Granny would not falsely believe that the cat eats the food every night.

Both solutions require an appeal to "methods of belief formation", "belief-forming processes", "bases of belief" or "evidence". The second is the most explicit, so let us formulate it. Let m_p be the proposition that the subject has formed a belief that p in some way w, where w is the way in which the subject actually formed it:[7]

Method-relative basic safety Kp only if $\Box_{\mathsf{C}}((\mathrm{B}p \wedge m_p) \rightarrow p)$.

I call the amendment "method-relative *basic* safety" because just like basic safety proper, the revised condition only looks at situations in which one believes the same proposition as in the actual one. So the revision would face the three previous worries for basic safety.

Now that the safety condition is saddled with a reference to methods, safety faces the generality problem: if methods are individuated too narrowly, any true belief is safe, if they are individuated to broadly, any belief is unsafe.[8] Of course, one can appeal to methods while remaining silent as to how they are individuated —one could use the intuitions about knowledge ascriptions as a guide to individuate them— but then the account is threatened of being trivially true. I have no solution to the generality problem on offer; here I am just pointing out that safety condition is not a way to dodge it.

6 Generalized safety

Since we are saddled with methods anyway, the natural suggestion is to use them in order to deal with the other worries. Here is the rough idea, for each of them:

1. *Fake-barn cases*. The relevant possibility of error in fake barn cases is either (a) there being a fake-barn at the current location of the barn, and S having a false belief about the fake-barn, or (b) S looking in a slightly different area of the field and having a false belief about a fake barn there, instead of a true belief about the real barn he is looking at. We may say that both beliefs are beliefs that would be formed by the very same method with which the actual one is formed. So the subject's actual belief is unsafe, in the sense that it has been formed by a method which could easily have yielded false beliefs.

[7] That would need to be amended to deal with embedded uses of "know", but let us ignore those issues here.

[8] The generality problem has been first formulated as a problem for reliabilism, by Goldman Goldman [1979] and by Conee and Feldman [Conee and Feldman, 1998].

2. *Borderline ignorance cases*. The relevant possibility of error in borderline ignorance cases is that one could have had the same disposition to apply "small number" (and a corresponding concept) even though the phrase had a slightly different meaning (or the concept would have been slightly different). If we consider that disposition as the subject's method, we can say that the slightly different belief would have been formed by the same method but be false. So the subject's actual belief is unsafe as above.

3. *Lucky guesses and bad methods cases involving necessary truths*. Consider lucky guessing whether a number is prime as the method used by Bob. The same method could easily have yielded a false belief about a slightly different number; so the original belief is unsafe in the same that it has been produced by a method that could have produced a false belief.

Now the general suggestion is clear enough. The basic safety condition on a belief that p looks at counterfactual situations in which one also has a belief about p. The focus on alternative situations involving belief in the same proposition proved too restrictive (as fake barns, vague predicates and lucky guesses about necessary truths show) and too broad (as evidential luck cases show). A proper generalization of safety would look instead at all and only beliefs produced by the same method. Thus:

Generalized safety (informal) S's belief that p formed by method m is safe iff S could not easily have formed a false belief by the same method.

A formal version of the condition requires a bit of unorthodox formalism. We need to refer to beliefs in order to say things like "the method with which such-and-such belief has been formed". Here is one way to do it:

Generalized safety b is safe iff $b\mathrm{B}m \wedge \Box_{\mathrm{C}} \forall b' (b'\mathrm{B}m \rightarrow \mathrm{T}b')$

where b refers to a particular beliefs, b' is a variable over beliefs, $b\mathrm{B}m$ holds iff belief b is based on method m, and $\mathrm{T}b$ holds iff belief b is true.

7 Discussion of generalized safety

The generalized safety condition comes very close to a version of reliabilism. Goldman's reliabilist account of knowledge [Goldman, 1986] requires "global" and "local" reliability for a true belief to be knowledge. The global condition is that the belief is formed by a process which generally yields true beliefs. The local condition is that the process necessitates a true belief in the situation at hand. Goldman does not formulate local reliability in a more detailed way, but the idea seems to be the same as generalized safety as we defined it above.

Virtue or reliabilist conditions have been put forward as a complement to safety. Safety is thought to be insufficient because one's belief can be safe (in the basic sense) without one's belief be formed in a proper, knowledge-conferring way. But given the new, generalized version of safety, it is unclear that such a complement is needed. So here as well, it is unclear whether the generalized safety account differs from some virtue accounts of knowledge.

That opens the possibility that generalized safety is sufficient for knowledge. Yet there are objections to that, notably Pritchard's Temp case.

References

Earl Conee and Richard Feldman. The generality problem for reliabilism. *Philosophical Studies*, 89(1):1–29, January 1998. doi: http://dx.doi.org/10.1023/A:1004243308503. URL http://dx.doi.org/10.1023/A:1004243308503.

Alvin I. Goldman. *Epistemology and Cognition*. Harvard University Press, 1986.

Alvin I. Goldman. What is justified belief? In G. S. Pappas, editor, *Justification and Knowledge*. Reidel, Dordrecht, 1979.

Robert Nozick. *Philosophical Explanations*. Belknap Press, Cambridge, Mass., 1984.

Ernest Sosa. How to defeat opposition to moore. In James Tomberlin, editor, *Philosophical Perspectives*, volume 13, pages 141–53. 1999. doi: http://dx.doi.org/10.1111/0029-4624.33.s13.7. URL http://dx.doi.org/10.1111/0029-4624.33.s13.7.

Peter Unger. An analysis of factual knowledge. *Journal of Philosophy*, 65(6):157–170, 1968. URL http://links.jstor.org/sici?sici=0022-362X%2819680321%2965%3A6%3C157%3AAAOFK%3E2.0.CO%3