Reliabilism – Modal, Probabilistic or Contextualist[1]

Peter Baumann

Swarthmore College

Summary

This paper discusses two versions of reliabilism: modal and probabilistic reliabilism. Modal reliabilism faces the problem of the missing closeness metric for possible worlds while probalistic reliabilism faces the problem of the relevant reference class. Despite the severity of these problems, reliabilism is still very plausible (also for independent reasons). I propose to stick with reliabilism, propose a contextualist (or, alternatively, harmlessly relativist) solution to the above problems and suggest that probabilistic reliabilism has the advantage over modal reliabilism.

Reliabilism about knowledge has it that knowledge is reliable true belief, that is, true belief which has been acquired in a reliable way. We do not have to take this as a reductive definition of knowledge. It seems hopeless to try to give a reductive definition of any ordinary (and philosophically interesting) concept, like the concept of knowledge. Let us rather see it as an indication of a conceptually necessary condition of knowledge:

(KR) Necessarily, if S knows that p, then S acquired the belief that p in a reliable way.

1 Published in *Grazer Philosophische Studien* 79, 2009, 77-89.

The general idea of knowledge reliabilism strikes me as being very plausible and Alvin Goldman deserves major credit for developing hints by Frank Ramsey into a full-blown theory (cf. Ramsey 1990a, 91-94, 1990b, 110; Goldman 1992a, 1992b, 1986, 1988, 2008; cf. also Armstrong 1973, Dretske 1981 and Nozick 1981). The basic idea can be formulated in such generality that it covers both internalism and externalism about knowledge. It can also be applied to other epistemic notions, like the concept of epistemic justification, or to non-epistemic notions, like the concept of moral character. In the following I restrict myself to general process reliabilism and will not go into the different forms of "local" reliabilism (sensitivity, safety, relevant alternatives account, etc.; cf. Becker 2007 for a combination of sensitivity with Goldman's account).

So, reliabilism looks great in general but things become tricky when we look at the details. What exactly is meant by "ways of belief acquisition" (or "processes" as I will simply call them here, - covering both what Goldman 1986, 93 calls "processes" and "methods") and how should they be individuated? And what exactly is meant by "reliable"?

Let me start with the last question. A first, rough answer says that a process is reliable just in case it tends "to produce beliefs that are true rather than false" (Goldman 1992b, 113). Here is one interpretation of this remark: A process P is reliable just in case the ratio t/ f of the number t of true beliefs to the number f of false beliefs resulting from the process is above a certain value r.[2] The value of "r" need not be a precise one and might well vary with context. The general idea here is that

2 The ratio t/f is taken to approximate a limit. - I am using the term "result" in a causal sense here. - What if P also produces beliefs without truth-values? What if we give up the idea of bivalence? What if the process does not produce any beliefs? We can disregard these questions here because the main points below do not depend on these issues.

(R) P is reliable iff t/ f > r.[3]

(R) only holds for what Goldman calls "belief-independent processes" (cf. 1992b, 117); we don't have to go into the complications having to do with belief-dependent cases here (where we would have to add the condition that the basing beliefs are true). I also skip the necessity operator for the sake of simplicity here as well as in the formulations of the following principles. We certainly have to add certain whistles and bells, like the restriction to normal (human) subjects and to normal circumstances for the running of the process. These additional constraints are important and not without problems; however, since not much depends on this here, I will disregard them here for the sake of simplicity.

Interesting questions arise: What do we mean by "beliefs resulting from the process"? All such beliefs, past, present and future, ever acquired by any subject? How can we refer to all such beliefs in our reliability judgments when we only have a very restricted sample of beliefs available? Are we justified in assuming that our sample is representative? I don't want to further go into this because there is an even more interesting and more worrisome problem: Should we restrict t/ f to the actual world? After dealing with this question (sections 1-2), I will discuss a non-modal, probabilistic alternative (sections 3-4).

---

3 I am assuming – also for the sake of simplicity – that all beliefs are created equal in the sense that the truth of some beliefs does not contribute more to the reliability of the process than the truth of some other beliefs. I am also assuming that any unequal distribution of true and false beliefs is due to the nature of the process running.

1. Modal Reliability

Goldman has proposed a negative answer to the last question: According to him, the notion of reliability is a counterfactual, modal one (cf. 1992a, 1986, 48, 107, 1988, 61-63; cf. Becker 2007, 11-12, 32, 89-92 in support but cf. also critically McGinn 1984, 537-539). We thus have to modify (R):

(M1) P is reliable in the actual world iff in the actual world (but cf. Goldman 1988, 61) and in some suitably qualified possible worlds $t/f > r$.

It seems obvious that a process can be reliable even if $t/f < r$ (or $\leq r$) in some worlds. It would be way too much to ask for if the ratio had to be above r in all possible worlds. Human vision can be reliable even if there are strange worlds in which it is useless.

This raises a difficult question: In which worlds does t/ f have to be above r in order to give us reliability for P? In 1986 Goldman proposed to restrict the set of possible worlds relevant here to what he called “normal worlds": worlds which fit our general beliefs about the actual world (cf. 107). This gives us

(M2) P is reliable in the actual world iff in normal worlds $t/f > r$.

There are several problems with this proposal – as Goldman later (cf. 1988, 62; cf. also Becker 2007, 33-34) acknowledged: Which beliefs held by whom fix the set of normal worlds? Goldman thus gave up the restriction to modal worlds but stuck with the idea that the notion of reliability is a modal notion.

How else could one restrict the set of possible worlds? One major problem with the idea of normal worlds is that it relativizes the set of the relevant possible worlds to our (true or false) beliefs about the actual world. What if all or most of these beliefs are false? Call a world "close" just in case it is similar in the relevant respects (what ever those are) to the actual world. Should we really say that a process is reliable even if in close but anormal worlds t/ f < r (or ≤ r)? It seems to make much more sense to say instead that

(M3) P is reliable in the actual world iff in close (normal or anormal) worlds (including the actual world) t/ f > r.

Again, here and elsewhere we don't have to take the idea of giving a full definition too seriously. I am assuming that this comes "close" to what Goldman agrees to nowadays. Let me therefore go into some problems I see with (M3) and then move on to an alternative, non-modal, account of reliability.

2. Problems with Closeness

The main problem is, of course, this: What determines whether a possible world is close to the actual world? Is there something – some matter of fact – which determines closeness? It is amazing that not much work has been done so far on this, crucial, question.

One quick response would be to say that the further characterization of the process will determine the set of close worlds. Take vision as an example and consider a

world which is so radically different from the actual world that we would not even say that vision exists in that world. The laws of optics might be so different that we would not want to call anything subjects are involved with in that world "vision". We thus have to restrict ourselves to worlds in which vision exists, and those worlds are the close ones. The problem with this kind of response is that it is too quick. Why should there not be remote worlds in which vision exists? Think of a world in which people are often envatted for some time. Or why should there not be close worlds in which vision does not exist? Such a world might not be epistemologically close but could still be metaphysically close. Or would we have to "define" closeness in epistemological terms? But why? And even if we did and had to: there could still be "close" (in that sense) worlds in which vision would not give us the right ratio of true and false beliefs. Think of a possible world in which people are often very absentminded and confused. The existence of such a world should, however, not make us deny the reliability of "our" vision (except if close worlds are only worlds where vision gives a high ratio of true beliefs - but such a stipulation would trivialize the point made).

So, the question remains wide open: What makes a possible world close to the actual world? Following Lewis (1973, 48-52, 66-67, 1986, 20-27), most people would explain closeness in terms of similarity. This is fine as it stands but it does not solve our problem. Everything is similar in some respect to everything else – so, which similarities count when it comes to closeness of worlds? Suppose that in the actual world I am not a brain in a vat, have never been one and will never be one. Compare the actual world to one possible world (WEIRD) in which I am not envatted but in which the laws of nature are very different from the actual ones; compare it also to another possible world (VAT) in which the laws of nature are the same as in the actual world

but in which I am envatted from time to time (cf. my 2005, 232-237 as well as Neta 2003, 16, fn.51 and Grobler 2001, 293). Is VAT closer to the actual world than WEIRD? Ask an epistemologist and you get one answer, ask a physicist (or even a metaphysician) and you get a different answer. Lewis 1979 tried to give criteria dealing with such cases but without much success, I think (cf. also Fine 1975, 451-458; Jackson 1977, 4-8; Slote 1978, 20-25; Bowie 1979; Heller 1999, 116; I cannot go into this further here). Interestingly, Lewis himself conceded that similarity and thus the closeness of worlds is context-dependent (cf. 1973, 50-52, 66-67, 91-95; cf. also Williams 1996 and Heller 1999, 505-507).

All this suggests that there simply is no such thing as the one and only one closeness (or remoteness) metric for possible worlds (no matter whether closeness is spelled out in terms of similarity or in other ways). At least, I think we have no good reason that there is such a thing. If this is correct, then there can be two (or more) different closeness rankings Ra and Rb such that one and the same process P comes out as being reliable, given Ra, but as not reliable, given Rb. If the world in which I am sometimes envatted is to count as close, then my vision might not count as being reliable while it could come out as reliable if that world was to count as a remote one. In a similar way, the answer to the question whether S knows that p will depend on the chosen closeness ranking.

How bad is this consequence? It depends. Our principle

(M3) P is reliable in the actual world iff in close (normal or anormal) worlds (including the actual world) t/ f > r.

would require either a further relativization

(M4) P is reliable in the actual world (given some closeness ranking) iff in worlds (including the actual world) which are close to the actual world[4] according to that ranking t/ f > r

or a contextualist interpretation according to which the truth value of the right-hand side of (M3) depends and varies with the closeness ranking used and presupposed by the attributor of reliability. In a similar way, knowledge attributions will have to be relativized or contextualized. It is not clear whether Goldman would want to have any of this. Should we rather give up knowledge reliabilism then?

3. Probabilistic Reliability

But why think of reliability as modal? Why not explain the notion probabilistically (where the concept of probability is not a modal one)? This approach has, I think, certain advantages. But let me start with a sketch of probabilistic reliability (cf. for this kind of approach: Kvart 2006).

The basic idea is to explain reliability in terms of the conditional probability (Pr) to acquire a true belief (T) as a result of a process aiming to settle a given question (I skip reference to the actual world from now on):

(P1) P is reliable iff Pr (T/ P happens) > s.

4 The actual world is close to itself.

Here and in the following further developments of the principle, "P" refers to a type (not token) process; for the sake of simplicity, I also skip the quantifiers (it being obvious what the full and cumbersome principles would look like). The value of "s" needs to be high enough (it need not be a precise one, though, and might well vary with context).[5] We can assume that it is unrealistic to expect that s = 1 for any process. Furthermore, in normal cases of reliable processes we can expect that s > .5.

(P1) isn't the complete story yet because the probability of a true belief might not in any way be due to the process. I want to include the idea of efficacy of the process in the idea of reliability of a process. It should make a difference whether – everything else remaining the same – P happens or does not happen. Let us therefore modify (P1) in the following way:

(P2) P is reliable iff

(a) Pr (T/ P happens) > s, and

(b) Pr (T/ P does not happen) ≤ s.

However, the additional clause (P2b) seems too strong: A process can still be reliable even if it raises the probability of a true belief just a bit from a value which is above s. It therefore seems more appropriate to say that

(P3) P is reliable iff

5 Again, we can disregard cases where the process results neither in a true nor in a false belief. – Similar to the modal principles above, (P1) and related principles only hold for belief-independent processes. We also have to restrict ourselves to normal subjects and normal circumstances.

(a) Pr (T/ P happens) > s, and

(b) Pr (T/ P happens) > Pr (T/ P does not happen).

4. Reference Class Problems

Now, how does (P4) (or similar probabilistic principles) fare as an explanation of the notion of reliability? There is a famous and notorious problem for reliabilism, the so-called generality problem (cf., e.g., Feldman 1985, 160-162, Alston 1995 and Goldman 1992b, 115-116, 1986, 49-51). I think it is not just one amongst several problems of the theory. Rather, it can be generalized in such a way that some very basic questions about the adequacy of a probabilistic account of reliability arise. So, what is the problem?

The problem is simply how to individuate processes of belief acquisition. One certainly does not want to be so specific that only one token of the process type exists. In that case, we would only have a reliability of 1 or of 0 but nothing in between. It would also not be advisable to individuate processes extremely broadly: The process of using one's cognitive apparatus does not seem to have some definite reliability. So, the correct individuation of a process is neither too narrow nor too broad. It should lie in the middle – but where exactly?

This is such a huge problem because even if we ignore the extremes it is still (at least very often) possible to find two (or more) different ways of individuating the process such that according to one the process is reliable while according to the other it isn't. Consider this case. Julie is looking at the sky and notices an airplane; she can even see that it is an Air France one. Just looking at the sky is not a reliable way of finding out what kind of airplane is flying by. But looking at the sky under today's very special

visibility conditions and after having taking eyedrops is a reliable method or process. And Julie has just done that. However, she is also suffering from a particular kind of headache which makes object recognition very difficult. Looking at the sky with that kind of headache, even under today's very special visibility conditions and after having taken eyedrops is not a reliable method or process. So, does Julie have a reliable true belief (or knowledge) that an Air France plane is passing by?

This depends on whether there is the one and only one right way of picking out and describing the process. If not, then there is no fact of the matter concerning reliability. To many, this would seem like a very bitter theoretical pill to swallow. And it has proven extremely difficult to find a solution to the generality problem.

Goldman 1986, 50 proposes that the narrowest type of process which is causally operative in the production of the belief is the relevant one. However, what are we going to say if there is no strict causal relation between the process and its outcome? What if there is only a probabilistic one? How should we choose between a broader process with higher probabilistic correlation and a more narrow process with lower probabilitistic correlation? Also: Why should we go with the narrowest such process? I will come back to this and related problems below.

At this point, I do not want to go much more into the generality problem because it is just one special case of a much broader problem: the problem of the relevant reference class (cf., e.g, Gillies 2000, 119ff.). Here is an example. Mary is graduating from her university. What is the probability that she will find a job within the first three months after graduation? Statistics have it that 65% of female university graduates do find a job within three months after graduation. However, only 55% of graduates from Mary's university are so successful. Fortunately, Mary is from an upper class

background – and 90% of upper class graduates don't have to wait longer than 3 months for their first job. However, Mary had a child in her first term and young mothers struggle to find jobs after graduation. And so on. What then is the relevant reference class into which Mary belongs and which determines her job chances? To get back to cognitive processes: A particular type of process P is the relevant one just in case the token process belongs into the relevant reference class of all processes of type P. Now, our more general reference class problem remains even if we could solve the generality problem. Why?

Suppose we don't worry what the relevant process is and simply try to determine its reliability. Consider the good old fake barn case (cf. Goldman 1992a, 86). Ernie finds himself in front of a real barn. He looks at it and acquires the true belief that there is a barn in front of him. Is, say, looking at an object of that size from that distance and under these conditions (let us just call this "looking") reliable? What if the following is also the case: All the other barns on the farm are fake, no other farm in the village has fake barns but all other villages in the county are full of them? Again, is looking reliable? The answer to this question depends on what the relevant spatial reference class is: the area around Ernie and the barn he is looking at, the farm, the village, or the county? I don't see how there could be a matter of fact which determines exactly one relevant spatial reference class. But couldn't we use probabilities to solve the problem? There is a certain probability, one might want to say, that Ernie will only ever look at this one barn, and another probability that he will never travel through the county, etc. However, the same questions can be raised with respect to these probabilities: What are their relevant reference classes? A regress of the reference class is lurking here.

Similar questions can be raised about temporal reference classes: with respect to which temporal intervals ought we to judge the reliability of the process? Suppose that Ernie's reliability varies with the time of the day, day of the week, season, etc., and we get a similar problem. Again, it is very doubtful whether there is a relevant reference class.

One general strategy to solve the reference class problem is to go with the narrowest probabilistically relevant reference class. In our temporal case above this would be the time of looking at the barn (let us assume that there is an uncontroversial beginning and end of the process and that probabilities don't change during the process). However, there are several problems with this strategy. One has to do with the fact that very often the narrowest reference class will have just one element. One difficulty here is that we might not have statistical information about single cases or about unique instantiations of a given cluster of properties. A deeper problem has to do with the question whether talk about probabilities in such a single case are meaningful at all. Even if we set these worries aside, there would still be the question why narrowness should matter in the first place? It is true that Ernie looked at the barn at time t but this does not entail that time t is the relevant time for the determination of the reliability of his vision. Much more could and should be said about this but I can leave it at that here.

Space and time are just two aspects. More aspects could be added. Let us also not forget that the same kind of problem arises with respect to the question what the relevant process or method used was. There is thus a whole bunch of reference class problems. The prospects of solving the reference class problem - in the sense of identifying a unique relevant reference class - seem dim (cf., e.g., Fetzer 1977; Hájek

2007). Even if one does not want to go that far and deny that there is a solution to the problem one would still have to admit that we just don't know what determines relevant reference classes. Given that our judgements about reliability and knowledge depend on this, this is still bad or interesting enough. What does all this imply for the notion of reliability?

As long as there is no variation of the probabilities along the relevant dimension (space, time, etc.): not much. However, we cannot make the assumption that this will always or even very often be the case. There will be at least some cases – and not too few – where there is such a variation of the probabilities. And in those cases, we will have no unique answer to the question whether the person has acquired her belief in a reliable way. If one is a reliabilist, one will therefore also have to conclude that at least in some cases there is no unique answer to the question whether S knows that p.

Again, the question is: How bad is that? And again, it depends. Our simple principle

(P1) P is reliable iff Pr (T/ P happens) > s

would require relativization to reference classes:

(P5) P is reliable with respect to a given set of reference classes RC iff Pr (T/ P happens under the conditions determined by RC) > s.

One might object that (P5) only gives us part of the story because we take the method or process as given and thus unaffected by the indeterminacy of the reference class.

However, this need not be a problem. Let "P" be a description of the method which is "meagre" enough so as not to allow for variations in the relevant probabilities. Everything else can then be subsumed under "circumstances" under which the process takes place. This is acceptable because nothing forces us to distinguish between process and circumstances in any particular way. If we go with a meagre enough description of the process, we will get a useful version of (P5).

An alternative would be to go contextualist and argue that the truth value of the right-hand side of (P1) and its kind depends on and varies with the reference classes chosen by the speaker. Indirectly, this will, of course, also lead to a relativism or contextualism about "knowledge", given reliabilism. Again, I wonder what Goldman would say.

One option would be to give up reliabilism in order to avoid all that. However, I think that there is no strong enough reason to do that. Reliabilism has a lot of independent plausibility. And as far as I am concerned, contextualism does not look like a bad option at all.

5. Conclusion

Finally, what about the alternative between modal interpretations and probabilistic interpretations of "reliability"? Aren't they more or less on a par, at least with respect to the issues discussed here? I don't think so. I think there are clear advantages on the side of the probabilistic version. Let me quickly mention two. First, closeness rankings of possible worlds seem restricted to ordinal rankings while the apparatus of probability theory can capture more than that and represent relations between differences of

probabilities. Second, probability theory is closer to home if you're a naturalist than modal logic. The natural sciences are happy to use probability theory but seem to have little use for modal notions. I would therefore propose three things (in the light of all of the above); stick with reliabilism, go for a probabilistic version of it, and accept the contextualist implications of all that. How happy Alvin Goldman would be with that, I don't know.

References

Alston, William P. 1995, How to Think about Reliability, in: Philosophical Topics 23, 1-29.

Armstrong, David M. 1973, Belief, Truth and Knowledge, Cambridge: Cambridge University Press.

Baumann, Peter 2005, Varieties of Contextualism: Standards and Descriptions, in: Grazer Philosophische Studien 69, 229-245.

Becker, Kelly 2007, Epistemology Modalized, New York & London: Routledge.

Bowie, G. Lee 1979, The Similiarity Approach to Counterfactuals, in: Noûs 13, 477-498.

Dretske, Fred I. 1981, Knowledge and the Flow of Information, Cambridge/MA: MIT Press.

Feldman, Richard 1985, Reliability and Justification, in: The Monist 68, 159-174.

Fetzer, James H. 1977, Reichenbach, Reference Classes, and Single Case Probabilities, in: Synthese 34, 185-217.

Fine, Kit 1975, Critical Notice [of Lewis 1973], in: Mind 84, 451-458.

Gillies, Donald 2000, Philosophical Theories of Probability, London: Routledge 2000.

Goldman, Alvin I. 1992a, Discrimination and Perceptual Knowledge, in: Alvin I. Goldman, Liaisons. Philosophy Meets the Cognitive and Social Sciences, Cambridge/MA & London: MIT Press, 85-103.

Goldman, Alvin I. 1992b, What Is Justified Belief?, in: Alvin I. Goldman, Liaisons. Philosophy Meets the Cognitive and Social Sciences, Cambridge/MA & London: MIT Press, 105-126.

Goldman, Alvin I. 1986, Epistemology and Cognition, Cambridge/MA & London: Harvard University Press.

Goldman, Alvin I. 1988., Strong and Weak Justification, in: Philosophical Pespectives 2, 51-69.

Goldman, Alvin I. 2008, Reliabilism, in: The Stanford Encyclopedia of Philosophy (Spring 2008 Edition; ed.: Edward N. Zalta), URL = http://plato.stanford.edu/entries/reliabilism/

Grobler, Adam 2001, Truth, Knowledge, and Presupposition, Logique-et-Analyse 44 (173-174-175): 291-305.

Hájek, Alan 2007, The Reference Class Problem is Your Problem too, in: Synthese 156, 563-585.

Heller, Mark 1999, The Proper Role for Contextualism in an Anti-Luck Epistemology, in: Philosophical Perspectives 13, 115-130.

Jackson, Frank 1977, A Causal Theory of Counterfactuals, in: Australasian Journal of Philosophy 55, 3-21.

Kvart, Igal 2006, A Probabilistic Theory of Knowledge, in: Philosophy and Phenomenological Research 72, 1-43.

Lewis, David 1973, Counterfactuals, Oxford: Blackwell.

Lewis, David 1979, Counterfactual Dependence and Time's Arrow, in: Noûs 13, 455-476.

Lewis, David 1986, On the Plurality of Worlds, Oxford: Blackwell.

McGinn, Colin 1984, The Concept of Knowledge, in: Peter French/ Theodore Uehling, Jr./ Howard Wettstein (eds.), Midwest Studies in Philosophy 9 (Causation and Causal Theories), Minneapolis: University of Minnesota Press, 529-554.

Neta, Ram 2003, Contextualism and the Problem of the External World, in: Philosophy and Phenomenological Research 66, 1-31.

Nozick, Robert 1981, Philosophical Explanations, Cambridge/MA: Harvard University Press 1981.

Ramsey, Frank Plumpton 1990a, Truth and Probability, in: Frank Plumpton Ramsey, Philosophical Papers (ed.: D.H. Mellor), Cambridge: Cambridge University Press, 52-94.

Ramsey, Frank Plumpton 1990b, Knowledge, in: Frank Plumpton Ramsey, Philosophical Papers (ed.: D.H. Mellor), Cambridge: Cambridge University Press, 110-111.

Slote, Michael A. 1978, Time in Counterfactuals, in: Philosophical Review 87, 3-27.

Williams, Michael 1996, Unnatural Doubts. Epistemological Realism and the Basis of Scepticism, Princeton: Princeton University Press.