Evidentialism and the Numbers Game[1]

Introduction

This paper poses a puzzle concerning a broadly held view about normative reasons for belief: *evidentialism*. Evidentialism is the highly intuitive view that the only normative reasons for belief are evidential reasons. I shall argue that in certain circumstances, evidentialism is unable to generate the correct reasons for belief; these reasons can only be provided by other kinds of epistemic reasons apart from evidential ones. I am not arguing that reasons in ordinary cases for belief are non-evidential, but that evidentialism is too narrow an account of normative reasons for belief to serve as a complete theory of epistemic reasons.[2]

Reasons for belief

It will be helpful to begin by giving a definition for an *evidential reason for belief*.

1. Fact *f* is an evidential reason for agent *a* to believe *p*, if and only if *f* is a reason to believe *p* and *f* is evidence for *p*.[3]

With a definition in hand for evidential reasons for belief, we can now define a *non-evidential reason for belief*.

---

[1] This paper has benefited significantly from a number of comments. I owe particular thanks to Bill Brewer, John Broome, Roger Crisp, Sven Nyholm, Jeff Speaks, Wlodek Rabinowicz, and two anonymous referees for *Theoria*. I have also benefited from comments given at presentations of this work in a slightly different form at the Oxford Moral Philosophy Seminar and in much the present form at the Canadian Society for Epistemology's International Symposium on Rationality in Contemporary Epistemology.

[2] For a sample of contrasting views, see: Adler (2002), Feldman and Conee (1985), Railton (1994), and Shah (2006).

[3] One might wonder why it is necessary to speak of reasons for belief when one has already said that there is evidence for a belief, as it might seem that there is something internal to the concept of evidence that contains the appropriate concept of a reason. That evidence is analytically related to reasons for belief is not an uncontested view, c.f. Stich (1993). The schema used here allows logical space for discussion of non-evidential reasons for belief.

2. Fact *f* is a non-evidential reason for agent *a* to believe *p*, if and only if *f* is a reason to believe *p* and *f* is not evidence for *p*.

Evidentialism is the view that all normative reasons for belief must meet the criteria in 1. I do not propose to give a definition for evidence, as the details of differing theories of evidence do not matter to the argument here. For the sake of convenience, we may treat transmission of evidential warrant as a special case of evidence for purposes of evidentialism.

Self-affecting beliefs and evidentialism

Evidentialism faces a problem when there is, or appears to be, a normative reason to believe something, but there is no evidence for it. I want to focus on one particular class of cases in which evidentialism fails to provide reasons, where it looks as though it ought to: self-affecting beliefs, i.e. beliefs, the possession of which determines whether those beliefs are true. I shall use an example, which I shall call 'the numbers game'.

Alice is hooked up to a computer via a helmet that is capable of reading her thoughts. This helmet accurately reads her beliefs and transmits them to a computer that is attached to a monitor. Alice is then given the following complete, and completely accurate, set of rules for a game of sorts: the numbers game.

When hooked up to the computer, Alice will be told that five seconds from now, a number will appear on the screen, which is presently blank. She is told that the number on the screen will be 16, if she forms no belief at all about what number will be on the screen. If, however, she forms any belief about what number will be on the screen, then a different number will appear, a number which is half the number that Alice believes will appear plus 1. So, if she believes the number will be 6, it will be 4, and if she believes it will be 10, then it will be 6, and if she believes it will be 5 it will be 3.5, etc. If Alice forms a new belief in the interval between her forming an initial belief and a number's appearing on the screen, then the computer performs the operation again and one-half plus one of the new number that Alice believes will appear in five seconds. Alice knows all these rules, and faces a problem because of it.

In all cases but one, whatever Alice believes will be wrong. If she believes the number will be 8, it will be 5. Alice knows that if she forms no belief at all the number will be 16. So, once she recognises that she has no belief about the matter, she will believe that the number will be 16, and then of course the number will be 9. Coming to believe that the number will be 9, Alice will then be able to apply the formula and realise that the number will be 5.5, and so on. Alice will forever be stuck with unstable beliefs, unless she happens upon the single stable number: 2

Another way of putting this problem is that any belief Alice has will be false, unless that belief is that the number on the screen will be 2. If Alice has any reason to believe anything at all about what number will appear on the screen, she ought to believe that the number will be 2. That is the only belief that will be true. The interesting question is what reason is there for Alice to believe that the number will be 2? I propose that there are three possible reasons for Alice to believe that it will be 2. The first reason is that if she does, she will be correct. The second reason is that she will cease to have unstable beliefs, and the third is that there is evidence that it will be 2.

Only the third option could serve as a reason for an evidentialist. I shall argue that the fact that I will be correct, if I believe a particular proposition, is not in general evidence for that proposition. And, I shall argue that the fact that my beliefs will be stable is not in general evidence for the belief in question.

Believing something because you will be correct if you believe it

The first claim is that the fact that she will be correct is a reason for Alice to believe that the number on the screen will be 2. One might argue along these lines that the fact that one will be right is a reason to believe something. Beliefs aim at truth, and reasons for belief are related to the truth of that belief. Alice's belief about the number will be true if and only if she believes that the number will be 2. Alice knows that she will be right if and only if she believes that it is 2. As her beliefs aim at truth, her beliefs will meet their aim if and only if she believes it is 2. From this we may reasonably infer that the fact that she will be right if and only if she believes that it will be 2 is a reason for Alice to believe that it will be 2.

This line of reasoning may be open to an objection. The objection is that it leads to an overly permissive form of externalism about reasons. That you will be right would be a reason for you to believe any fact, no matter whether you could have any epistemic contact with it or not. There would be reasons to believe facts about epistemically inaccessible parts of the universe, facts about the psychology of people of whose existence you are not aware, and so on.

This objection can be met by providing a weakened version of the 'because you will be right' justification. The weakened version would hold that *within the appropriate limits of epistemic accessibility*, the fact that you will be right if you believe something is a reason for you to believe it. Because the rules of the numbers game are fully spelled out for her, Alice knows that if she believes that the number will be 2, then it will be 2. A strong limit on epistemic accessibility would hold that the fact that you will be right is a reason to believe something only if you know that you will be right if you believe it. In the case of the numbers game, there would be a reason for Alice to believe that the number will be 2, because she knows that she will be right if she believes that the number will be 2. The knowledge constraint may be too restrictive, but some weaker form of the epistemic accessibility constraint on when the fact that you will be right counts as a reason to believe something is available.[4]

Even with an epistemic constraint added, this justification is not open to the evidentialist. In order for a fact to be a reason for a belief, according to evidentialism, it must stand in an evidential relation to the contents of that belief. By evidentialist lights, that she will be right if she believes it is a reason for Alice to believe that the number on the screen will be 2, only if the fact that she will be right if she believes it is evidence that the number will be 2. Consider a standard, if simple, account of evidence: e is evidence for p when the conditional probability of p given e is greater than the conditional probability of p given not e. A look at Alice's situation shows that if Alice abides by the constraints of evidentialism, the probability that the number on the screen will be 2, given that Alice will be right if she believes it, is no greater than the probability that the number on the screen will be 2, given that it is not the case that Alice will be right if she believes it. What is needed is evidence that Alice *will* believe that the number on the screen will

[4] A stimulating proposal concerning epistemic accessibility constraints is offered in Skorupski (2002).

be 2. So far, there is only a reason for Alice to believe that if she believes the number will be 2, then it will be 2.

Can evidence be generated that shows it is likely that Alice will believe that the number will be 2? This depends a good deal on Alice. If Alice is a strict rational evidentialist, then we must answer this question in the negative. Consider Alice's epistemic situation. Alice has evidence for the material conditional: If she believes it will be 2, then it will be 2. If she is responsive to all and only her evidential reasons, she will believe that if she believes the number will be 2, then it will be 2.[5] But now she needs some evidence to show that the antecedent of the conditional will be affirmed: that she will believe the number will be 2. It is not clear why an evidentialist ought to believe of herself that, in this circumstance, she will believe that the number will be 2. Were Alice disposed to believe things when she will be right, if she were to believe them, then that might serve as evidence that she will believe that the number will be 2. But, this does not appear to be a disposition that a good rational evidentialist would have.

A rational evidentialist will not have this disposition, because in general the fact that something will be true, if you believe it, is not evidence that it is true. Suppose, for example, that I am a reliable believer about my own future actions, because I know my own intentions and am good at carrying them out. That the fact that I believe I shall do something is a reason to believe that I shall do it. So, suppose that if I were to believe that I shall start the day by swimming in the North Sea, then I would be correct. This conditional is not evidence that the well insulated sea-lions off the East Coast of Fife will have me as a companion tomorrow morning, for I certainly do not intend to put so much as a toe in the icy water.

Before moving on to stability as a reason for belief, it is worth considering whether the fact that one will be correct, suitable epistemic accessibility constraints applied, is an epistemic reason for believing something. There are two issues that must be resolved in answering this question. The first is whether the fact that one will be correct, if one has a particular belief, is a reason of any sort to have that belief. The second is whether the reason, if indeed there is a reason at all, is an epistemic one.

[5] I am not committing the rational evidentialist to the strong view that all beliefs formed through non-evidential routes are irrational (they may, for example, be non-rational), but rather that the only rational path to belief formation is via recognizing evidential reasons.

As to the first issue, it seems to me that there is a good case for thinking that the fact that one will be right, if one believes something, is a good reason for believing it. Two considerations support this. The first is pragmatic. Although it is not always the case, for the most part, having correct beliefs is more useful than not. We can cook up cases in which there are incentives to have false beliefs, but insofar as beliefs are our mental map of the world, we can navigate more successfully when our beliefs are true rather than false.

The second, and independent, consideration is that because our beliefs in some sense aim at truth, there is something to be said in favour of having true beliefs. The evidentialist herself trades on a similar intuition. The explanation of why evidence for the contents of a belief provides a reason for that belief is that something's being evidence for a belief makes it more likely (at least in simple cases) that the belief is true. That one will be correct if one believes something is one point better: having a particular belief insures its own truth.

Of course, we have to be careful about making a fetish out of having true beliefs. Many true beliefs are not worth having, or at least not worth having occurently, and one could easily shoot for maximising the number of true beliefs that one possesses by studying tautologies or easy to remember, but absolutely pointless, trivia. However, nothing about saying that there is some reason to believe something because one will be correct, if one believes it, implies that one ought to go about maximising the number of true beliefs that one possesses; it merely says that there is some reason to believe things when believing them will result in one's being correct. A parallel point could be made about beliefs, the reasons for which are evidential.

The first consideration in favour of thinking that the fact that one will be correct, if one believes something, is a reason to believe it does not clearly suggest that that reason is an evidential one. I appealed in that explanation to pragmatic considerations, and perhaps that is enough to end talk of the resulting reasons' being epistemic.

The second consideration, however, suggests that there is an epistemic reason. I have made an appeal to belief's special relation to truth, just as the evidentialist would. If we are not begging the question by assuming that the only epistemic reasons for belief are evidential, then there is at least some grounds for claiming that the fact that one will be correct if one believes something (under suitable epistemic accessibility constraints) is a reason to do so.

Stability of beliefs as a reason for believing something

The second possible justification for believing that the number will be 2 is that one's beliefs will be unstable otherwise. The reason for Alice to believe that it will be 2 is that otherwise, she will have unstable beliefs during the time she is playing the numbers game. Neither of the two most plausible explanations of why there is a reason not to have unstable beliefs is inconsistent with the requirements of evidentialism.

The first explanation is that there is something disvaluable about having unstable beliefs. The disvalue most plausibly comes from pragmatic considerations. The constant adjustment of beliefs as is required in the unstable belief scenario takes up time and cognitive resources. Quite aside from any other benefit that might accrue from being right, it would be much more pleasant not to have to adjust constantly one's beliefs about what number will appear on the screen. Whatever plausibility we assign to this explanation— and these kind of pragmatic reasons for belief are highly controversial— the evidentialist must reject it, as the fact that having a belief would be a benefit to you is not evidence for that belief.

The second explanation of why there is a reason not to have unstable beliefs is that there is a sort of incoherence to having unstable beliefs. Aside from the belief that the number will be 2, any belief one might have about what number will appear on the screen will cause that belief to be false, and this is problematic. In the first instance, it is doubtful that it is even possible to hold beliefs that one knows are bound to be false. The difficulty with beliefs that cause themselves to be false, when the believer knows that this is the case, is that it is not clear that such beliefs could be held at all. There is something like a version of Moore's paradox at work: I believe that the number will be *x*, although I know that it will not be *x*, if believe I it will be.

This stability explanation of why there is a reason to believe that the number on the screen will be 2 is open to the evidentialist, but in too limited a way. The evidentialist *can* say that the fact that having a belief will cause that belief to be false is evidence that the belief will be false, and so the evidentialist can say that there is a reason not to have unstable beliefs. However, what the evidentialist *cannot* do is parlay the reason for not having unstable beliefs into a reason to believe the stable belief.

The evidentialist cannot do this, because there is still no evidence that the stable belief is, or will be, true. The only evidence that the content of the stable belief is true in this instance would be that Alice believes, or is likely to believe, that the number will be 2. However, the fact that there is a reason for Alice not to believe the unstable beliefs is not evidence that Alice will believe or is likely to believe the stable one. In the numbers game, if Alice is responsive to her reasons, and if there are only evidential reasons, then Alice will simply not believe any of the unstable beliefs. The fact that she will not believe any of the unstable beliefs is not evidence that she will believe the stable one. Without some sort of evidence that she will believe the stable belief, there is no evidential reason for Alice to believe that the number will be 2.

Returning to the claim at the start of the paper, we may ask of stability, too, whether it provides a non-evidential epistemic reason for belief. Here, the issue is somewhat more complicated than it was when considering whether that one will be correct, if one believes something, is an epistemic reason to believe it.

In the case of stability, unlike in the case of correctness, there is no obvious appeal to the special relationship between belief and truth. Stability, as I have set it up, is supposed to be a different reason from correctness alone. Stability instead offers itself as a putative reason for believing something because of independent cognitive considerations. Having unstable beliefs— especially unstable beliefs that occupy one's attention and concentration— is a waste of cognitive resources that might be used for other tasks.

I am inclined to think that the cognitive advantages, when all else is equal, provide a reason for believing the stable belief. But, I am unsure of whether to class this reason as epistemic or non-epistemic. Which classification is correct depends on whether cognitive virtues not directly connected with truth considerations are counted as properly epistemic virtues or as properly pragmatic, but epistemically related, virtues. If one leans towards the former, then stability can provide Alice with an epistemic but non-evidential reason for believing that the number will be 2. If not, stability provides a non-epistemic reason for Alice to believe that the number will be 2, if there are any non-epistemic reasons for belief.

In search of evidence

What evidentialists need is evidence that it is likely that Alice will believe that the number will be 2. The fact that Alice is likely to believe that the number will be 2 is evidence that the number will be 2.

To show that Alice is likely to believe that the number will be 2, the evidentialist might try to show, using Alice's responsiveness to her reasons as evidence, that Alice will believe, or is likely to believe, that the number will be 2.

One strategy that the evidentialist might draw on to do this is to grant that, although there are no non-evidential reasons for belief, there are non-evidential reasons for agents to cause themselves to believe something. On this view, any of the non-evidential reasons to believe discussed above could be taken as a reason for Alice to *cause herself* to believe that the number will be 2. Alice is responsive to her reasons, and so she is likely to cause herself to believe that the number will be 2. Causing herself to believe that the number will be 2 will result in her believing that the number will be 2, so Alice is likely to believe that the number will be 2. That Alice is likely to believe that the number will be 2 is evidence that the number will be 2. So, there appears to be an evidential reason for Alice to believe that the number will be 2.

There are two difficulties with this evidentialist solution to the puzzle about reasons for belief raised by the numbers game. The first difficulty is that it falls to the evidentialist to show that there is a reason to cause oneself to believe something, when there appears to be a non-evidential reason to believe it. It is not entirely clear that evidentialism can help itself to the assumption that putative non-evidential reasons for belief are, in fact, reasons for causing oneself to believe. This is a substantive view about reasons that requires some defence.

The second difficulty reinforces the first. By a slight modification of the rules of the numbers game, it is possible to make causing yourself to believe something always result in your having an unstable belief. In the modified numbers game, the computer to which Alice is attached is able to detect whether or not she caused herself to have the belief she has about the number (under whatever specification of 'causing herself' fills the requirement the evidentialists set for it). If she causes herself to have the belief about the number, then the normal rule of taking her belief, dividing it by 2, and adding 1 no longer applies. Instead, the computer will display the number that Alice caused herself to believe multiplied by 2 with 1 added to the product. So, if Alice causes herself to believe that the number will be 2, then the number displayed will be 5. Causing herself to believe something will ensure that Alice's beliefs are unstable.

Once the new rule has been added and Alice's beliefs will be unstable if she has caused herself to believe them, there is no longer any evidence that Alice will believe that the number will be 2, based on her being responsive to reasons. This is because there is no longer a reason for Alice to cause herself to believe that the number will be 2, as doing so will ensure that the number will not be 2. There is no reason for Alice to believe that she will cause herself to believe that the number will be 2, and thus there is no evidence that she will believe that it will be 2. The evidentialist cannot provide a reason for believing the stable belief in the modified numbers game, even under the generous assumption that he might be able to in the original numbers game. On the other hand, the position of the non-evidentialist is unaffected by the modification to the numbers game, as there are possible non-evidential reasons for Alice to believe that the number will be 2.

There is an alternative open to the evidentialist, however. Consider the following three claims:[6]

1. Alice has evidence that if she believes the number on the screen will be 2, then it will be 2.

2. Alice has evidence that if she believes *p*, then her belief that *p* will be based on evidence.

2*. Alice has evidence that if she believes *p*, then her belief that *p* will be based

[6] This line of argument was suggested to me both by an anonymous referee and by Jeff Speaks.

on *sufficient* evidence.

The first claim is true in virtue of being a consequence of the rules of the numbers game. Claims (2) and (2*) are versions of rational evidentialism, (2*) putting a stronger condition on rational belief than (2). Given these claims, the challenge for the rational evidentialist is to show that Alice has evidence, or perhaps sufficient evidence, that she will believe *p*. I have already argued that, because material conditionals do not constitute evidence for their consequents, Alice's belief that if she believes the number will be 2, then it will be 2, is not evidence that she will believe that the number will be 2.

An evidentialist could try to create evidence by asserting that Alice will be a rational evidentialist if:

R. Alice is disposed to believe *p*, if she knows that her belief that *p* will be based on evidence and that there will be no belief incompatible with *p* that is based on evidence.

or

R*. Alice is disposed to believe *p*, if she knows that her belief that *p* will be based on *sufficient* evidence and that there will be no belief incompatible with *p* that is based on *sufficient* evidence.

Combined with (1) and either (2) or (2*), the dispositions attributed to Alice in (R) or (R*) look sufficient for giving her the belief that she will believe that the number on the screen will be 2. Alice does know that if she believes *p*, her belief that *p* will be based on (sufficient) evidence. And, it is reasonable to think that there is no evidence (or not sufficient evidence) for any belief incompatible with *p*.

The difficulty with this line of argument is that there are good reasons to doubt that (R) and (R*) are plausible accounts of why a rational evidentialist is disposed to believe some particular proposition. Suppose Alice is a first rate mathematician. Appealing to (R) or (R*) as an explanation of why she comes to believe simple arithmetical propositions is quite odd. For example, consider how she would come to believe that 5 + 3 = 8 on the basis of (R) or (R*). Alice knows that if she believes that 5 + 3 = 8, then she will believe this because there is sufficient mathematical evidence. Furthermore, as simple arithmetical equations involving the addition of two integers are either clearly true or clearly false, there will be no (or at least vastly insufficient) evidence for any arithmetical proposition incompatible with 5 + 3 = 8. According to (R*), Alice will be disposed to believe that 5 + 3 = 8, because she knows that her belief will be based on sufficient evidence and that there will be no belief that conflicts with this proposition.

This explanation of why a rational evidentialist would be disposed to believe that 5 + 3 = 8 is forced and implausible. No doubt the most plausible explanation as to why a rational evidentialist would be disposed to believe that 5 + 3 = 8 is that she understands the rules of arithmetic and sees that they provide mathematical evidence that sum of 5 and 3 is 8. (R) and (R*) do not provide natural explanations in general of why a rational evidentialist would be disposed to believe something, and indeed the only reason I can see for introducing them is to account on an *ad hoc* basis for puzzles like the one that arises in the numbers game.

The fundamental difficulty one faces in searching for the kind of evidence that a rational evidentialist could believe is that to generate evidence, some sort of evidential circularity is likely to be required. The rational evidentialist must have evidence that she will believe that the number will be 2, in order for her to have evidence that the number will be 2. What would count as evidence that the number will be 2 is that there is evidence for the rational evidentialist that she will believe that the number will be 2.

Under ordinary circumstances, in order for the rational evidentialist to have evidence that she will believe that the number will be 2, she requires evidence that the number will be 2. And at this point, the rational evidentialist finds herself back where she started.

Conclusion

The numbers game represents a difficult problem for the evidentialist. There are at least two plausible considerations in favour of believing the stable belief: the fact that doing so is the only way to be correct and the fact that doing so is the only way to avoid having unstable beliefs. Neither of these types of reasons is available to the evidentialist. These considerations in favour of believing the stable belief intuitively look like good reasons for believing the stable belief, and evidentialism's inability to accommodate them is problematic. The modified version of the numbers game shows that even if there is some initial plausibility to the view that there is a reason to cause yourself to believe something when there appears to be a non-evidential reason to believe it, there will be situations in which there is clearly no reason to cause yourself to believe something on those grounds.

The arguments of this paper suggest that evidentialism is too narrow an account of epistemic normative reasons for belief. A complete theory of epistemic normative reasons for belief will have to have richer resources than those available to strict evidentialism, although this is not inconsistent with the view that in most ordinary cases, the normative reasons for belief are evidential ones.

References

ADLER, JONATHAN, *Belief's Own Ethics*. Bradford Books, 2002.

FELDMAN, RICHARD and CONEE, EARL, 'Evidentialism'. *Philosophical Studies* Vol. 48, 1985, pp. 15-34.

RAILTON, PETER, 'Truth, Reason, and the Regulation of Belief'. *Philosophical Issues* V, 1994.

SHAH, NISHI, 'A New Argument for Evidentialism'. *The Philosophical Quarterly* Vol. 56, 2006, pp. 481-498.

SKORUPSKI, JOHN, 'The Ontology of Reasons'. *Topoi* Vol. 21, pp. 113-124.

STICH, STEPHEN, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. MIT Press, 1993.