What Lottery Problem for Reliabilism?

Juan Comesaña

Abstract

It can often be heard in the hallways, and occasionally read in print, that reliabilism runs into special trouble regarding lottery cases. My main aim in this paper is to argue that this is not so. Nevertheless, lottery cases do force us to pay close attention to the relation between justification and probability.

1. Introduction

Reliabilism is one of the leading theories of epistemic justification that appeals to probabilistic considerations. Very roughly, reliabilism is the claim that the justification of a belief is a function of the truth-conduciveness of the process that created that belief. However, the process need not be *perfectly* reliable in order to issue in justified beliefs, and it is because of this reason that reliabilism must appeal to probabilistic considerations. Several issues have to be clarified before reliabilism can be evaluated: how exactly to identify the process of belief formation, the situation with respect to which the truth-conduciveness of the process has to be evaluated, and what it is for a process to be truth-conducive, among others. It is perhaps surprising that, of these issues, the one that has received the *least* attention in the literature is the question of what it means to say that a process of belief-formation is truth-conducive. There are hints here and there that the answer must have *something* to do with the probability of the proposition believed, but the issue has not received sustained scrutiny. The relationship between probability and reliability is the central topic of this paper.

In section 2 I present a schematic version of reliabilism that is able to deal with one of the main problems for reliabilism, the ‘generality problem.’ I present a version of reliabilism that is able to deal with the generality problem because otherwise it isn't clear what verdicts reliabilism gives regarding specific cases. This version of reliabilism is schematic, however, because it doesn't involve an explicit characterization of reliability. In section 3 I give an account of reliability in terms of truth-ratios in order to contrast it with an explicitly probabilistic account, developed in section 4. Finally, in section 5 I present a number of problems that lotteries raise for epistemic theories and argue that either they are not peculiar to reliabilism or they can be solved by appealing to the probabilistic account of reliability.

2. Reliabilism and the Generality Problem

Let us begin with a schematic formulation of reliabilism as a theory of epistemic justification:

> **Reliabilism**: A belief that *p* by *S* is justified if and only if that belief was produced by a reliable belief-forming process.

There is a serious problem that reliabilism must face: what has come to be known as 'the generality problem.'[1] Briefly put, the generality problem arises because we can only assess the reliability of a *type* of process, and any *token* process of belief formation belongs to indefinitely many types which differ widely in their reliability.[2] In order for reliabilism to have any consequences regarding particular beliefs, we first have to have a way of identifying which type of process is the relevant one. Although the main aim of this paper is not to solve the generality problem, I will briefly describe how I think it can best be dealt with, because the resulting reliability theory will be better equipped to deal with questions about probability.

Elsewhere I have argued that the correct solution to the generality problem involves incorporating the notion of *evidence* into reliabilism.[3] One theory that makes the notion of evidence the central one is Conee and Feldman's evidentialism:

> **Evidentialism**: Doxastic attitude *D* toward proposition *p* is epistemically justified for *S* at *t* if and only if adopting *D* toward *p* fits the evidence *S* has at *t*.[4]

Conee and Feldman do not wish to take the notion of evidence as a primitive: they explain that they think that evidence is constituted by (non-factive) mental states of the subject, such as justified beliefs that the subject has as well as experiences (and maybe apparent memories and 'intellectual seemings,' if there are any such things). But they do want to take the notion of what it is for a doxastic attitude toward a proposition to *fit* the evidence that a subject has as primitive. In other words, their theory explains what evidence is, but not what *good* evidence is.

The key in solving the generality problem lies in combining reliabilism with evidentialism. In a nutshell, the idea is that the justificatory status of a belief is a function of the evidence possessed by the subject (thus far evidentialists are right), and that the function in question is reliabilist in nature (and in this sense reliabilists are right too). So, applied to the generality problem, the guiding idea is that the relevant process to assess for reliability has the following form: *S's believing that p based on evidence e*, where 'p' is to be replaced by the proposition believed in each case and 'e' is to be replaced by the evidence on which the subject S bases his belief.[5]

With this solution to the generality problem in place, we can then formulate reliabilism as follows:

> **Evidentialist reliabilism**: *S*'s belief that *p* is *prima facie* justified if and only if *S* has evidence *e*, *S* bases his belief that *p* on *e*, and
>
> (i) either *e* contains no other beliefs of *S* and the type *S's believing that p based on e* is reliable; or
>
> (ii) *e* contains other beliefs of *S*, *S* is justified in having those beliefs, and the type *S's believing that p based on e* is reliable.[6]

This version of reliabilism doesn't face the generality problem: for any case of belief formation, it claims that the resulting belief is justified if and only if a certain specified type is reliable. The resulting reliabilist theory cannot be evaluated, though, until we know what it is for a specified belief-forming process type to be reliable. The rest of this paper deals with that question.

3. Reliability as a High Truth-Ratio

What is it for a certain process type to be reliable? That is, how should we fill in the gap in the following schema?

> **Reliability**: a type of the form *believing that p based on e* is reliable if and only if …

The answer must have *something* to do with the truth of the belief formed by that process. It cannot be, of course, that a process type is reliable if and only if it issues *only* in true beliefs, for in that case reliabilism would collapse justification with truth. Still, the process must be a *good guide* to the truth of the beliefs that it produces.

Informal explanations of the notion of reliability usually contain phrases such as 'a process is reliable just in case it generally leads to true beliefs.' Phrases such as this suggest that the reliability of a process is a matter of a high truth-ratio among its applications. Depending on whether we focus only on *actual* applications or also on *counterfactual* ones, we get as a result the following two specifications of the notion of reliability (for some suitable *r*):

> **Reliability as high actual truth-ratio**: a type of the form *believing that p based on e* is reliable if and only if the ratio of true to false beliefs in its actual applications is higher than *r*.

> **Reliability as high counterfactual truth-ratio**: a type of the form *believing that p based on e* is reliable if and only if the ratio of true to false beliefs in actual and close counterfactual applications is higher than *r*.

Reliability as high actual truth-ratio is both too narrow and too broad as a definition of reliability. It is too narrow because it is possible for there to be processes that confer justification even though the ratio of true to false beliefs in their actual applications is not very high—for instance, it is possible for there to be processes that we would judge as justifying their output beliefs even if, by chance, the few times that they were used they issued in false beliefs. Reliability as high actual truth-ratio is also too broad because the reverse situation is also possible: it is possible for there to be processes that we would judge as not justifying their output beliefs even if, by chance, the few times that they were used they issued in true beliefs.

For those reasons, reliability as counterfactual truth-ratio is preferable to reliability as actual truth-ratio. But the counterfactual account has its own problems. Given that we are not requiring perfect reliability, we must specify *which* counterfactual applications of the process are to be taken into account (which of those situations count as ‘close,’ to use a common terminology). Presumably, applications of the process that yield mostly false beliefs because applied in very unusual circumstances shouldn't affect its reliability. But the question of what circumstances count as usual and which don't is a vexed one. One attractive idea is that the situations that count as close to the actual case are those that are highly probable given the actual case. But if we are going to invoke probability at this stage in the deployment of a reliabilist theory, maybe we should explore a reliabilist theory that bypasses talk of closeness relations among possible worlds and is characterized instead directly in probabilistic terms.

4. Reliability as High Conditional Probability

Philosophers working on confirmation theory have studied different ways in which one can explain in probabilistic terms how (and to what extent) a scientific theory is confirmed by some evidence. In this section I will make use of some of these results in confirmation theory. However, there are important differences between the use to which probability theory will be put here and its use in confirmation theory, differences that, if ignored, may result in misguided objections to the resulting reliabilist theory. The main differences are three.

In the first place, a common interpretation of the probabilities in question in confirmation theory is as *subjective* degrees of belief of the subject in question. But if we are going to use probabilities to define reliability, the probabilities in question cannot be *subjective* probabilities, no matter whether we understand such notion in a descriptive or in a normative sense. Understood descriptively, the suggestion would be to tie the reliability of a process of belief formation to what the subject happens to believe about the relevant probabilities, no matter how crazy the subject's beliefs are. This certainly won't do. Understood normatively, the suggestion would be to tie the reliability of a process of belief formation to what a subject *ought* to believe about the relevant probabilities. Now, the *ought* in question is either going to be epistemic or of some other kind. If it is epistemic, then little progress has been made on the project of explaining justification in non-epistemic terms (something that, for better or worse, is a traditional reliabilist goal). If it is of some other kind, then, as with descriptive subjectivism, those probabilities have nothing to do with epistemic justification. If reliability is connected to probability, then, it is connected to some kind of *objective* probability.[7]

Second, when philosophers of science talk of probabilistic confirmation, likelihood ratios, etc., the information that, e.g., the evidence confirms a certain hypothesis over another to a certain degree is supposed to *be known* by the scientists, and it is through this knowledge alone that it can influence their attitudes towards a hypothesis. In other words, all (degrees of) belief in hypotheses are supposed to be inferentially arrived at. In using probability theory in the definition of reliability, on the other hand, we aim to clarify both inferentially *and* non-inferentially acquired belief (and, even within the class of inferentially acquired beliefs, we aim to cover cases where the inference in question is not probabilistic in nature). In keeping with the externalist character of reliabilism, what matters is what the probabilities are, not what the subjects think about the probabilities.

Third, and relatedly, the notion of *evidence* used in the version of reliabilism defended here is different from the notion of evidence used in philosophy of science. In philosophy of science, evidence is given by propositions like "the paper turned blue" or "Uranus's orbit has such-and-such properties." Those propositions, again, are supposed to be learned by the subject and then *used* as evidence in order to arrive at the degree to which the hypothesis is confirmed. In the case of the theory that we are considering, on the other hand, the evidence is given by propositions that describe mental states of the subject. Only *some* of those states will be other (justified) beliefs of the subject. In the basic cases, they will be, for instance, experiences that the subject has.

How *could*, then, reliability be connected to probability? The suggestion that immediately comes to mind, and the one that is discussed to a minimal extent in the literature, is that the reliability of a process is a matter of the conditional probability of the proposition believed given the evidence on which the belief is based:

> **Reliability as high conditional probability**: a type of the form *believing that p based on e* is reliable if and only if $Pr(p|e) > r$;

Where $Pr(A|B)$ denotes the conditional probability of A given B and r is some suitably selected number between 0 and 1.

Reliability as high conditional probability has been endorsed by William Alston. Alston's terminology is different from the one used here: instead of saying that a belief is justified if and only if it is based on evidence that is reliably connected to the truth of the proposition believed, Alston says that a belief is justified if and only if it is *adequately grounded*. Despite this terminological difference, the views are closely related. A *ground* is a mental state of the subject that he takes into account in forming the belief, and for it to be *adequate* '… it must be "truth conducive"; it must be sufficiently indicative of the truth of the belief it grounds. In other terms, the ground must be such that the *probability* of the belief's being true, given that ground, is very high. It is an objective probability that is in question here.'[8]

We have now introduced two different ways of thinking about reliability: in terms of the truth-ratios of the (actual or possible) applications of the process in question, and in terms of the conditional probability of the proposition on the evidence. It is now time to confront those two different conceptions of reliability with a problem that the theory is said to face.

5. Reliability and Lotteries

5.1 The Many Lottery Problems

Lotteries raise many interesting epistemological issues, but not all of them are directly relevant to reliabilism. I will present three lottery problems that are not particularly relevant to reliabilism, and a fourth one that is. Dealing with that fourth problem will be the main burden of the rest of this paper.

First, there is the issue of how to understand the relation between belief as an all-or-nothing notion on the one hand and degrees of belief on the other. Many philosophers think that there are two related doxastic phenomena: on the one hand, we can characterize our doxastic life in terms of whether we have one of three mental attitudes with respect to any proposition p—we can accept the proposition, reject it (i.e., believe its negation), or suspend judgment with respect to it (which, as a positive mental attitude which can be more or less justified, should be clearly distinguished from adopting no attitude whatsoever towards a proposition). On the other hand, we can characterize our mental life in terms of a much more fine-grained range of attitudes that we can take with respect to any proposition *p*—we can believe it to a very low degree, or to a low-ish degree, or to a high, or very high degree, etc. There is an initially attractive principle that links the ternary notion with the degree notion: *S* accepts that *p* iff *S* believes that *p* to degree *n* (for a suitably high *n*). Similar principles could be offered for linking rejection and suspension of judgment to degrees of belief. Another initially attractive principle is that if *S* accepts that *p* and *S* accepts that *q*, then *S* accepts that *p and q*. Lotteries present a problem for the conjunction of these two principles. For suppose that there is a fair lottery with a very high number of tickets. Then, I will believe to a very high degree, of each ticket, that it will lose. Therefore, given the first principle, I accept, of each ticket, that it will lose. But, given the second principle, I accept that all tickets will lose. But I certainly don't accept anything like that! A similar argument can be run in terms of rational acceptance.[9] It is clear that this problem raised by lotteries is not peculiar to reliabilism: every epistemological theory that wants to say something about the relation between the two ways of characterizing our doxastic life must face it.

Second, lotteries can be used to raise a particularly interesting closure problem. For I know that I won't have enough money to afford an African Safari for me and twenty of my friends next summer. I also know that if I won't have enough money, then the lottery ticket that I hold will lose. But I don't know that the lottery ticket that I hold will lose.[10] Obviously, the closure problem is not peculiar to reliabilism either.

Third, there is the problem of explaining exactly *why* we don't know that a lottery ticket will lose, when we base that belief merely on the very low probability that it has of winning.[11] It might seem as if reliabilism is a particularly salient target of this problem. For, according to reliabilism, what determines whether a belief is justified has something to do with its probability, and we can make the probability that the ticket will lose as high as we wish, and we will still think that it doesn't amount to knowledge. But what is, exactly, the problem? The main problem cannot be that reliabilism has the consequence that our lottery beliefs are justified, because that they *are* justified is a widely-held view.[12] Some may think that the problem lies in trying to give an account of knowledge as reliably produced, un-gettierized true belief. Lottery cases are not standard Gettier cases in that they don't involve a bad luck/good luck combination,[13] and so such an account would have the consequence that a subject who believes that she holds a losing ticket knows that she does. This result at least goes against current trends in epistemology, and so it might seem to be a problem. But it isn't. In standard lottery cases, all of this is assumed to be true: (i) that the subject is justified in believing that she holds a losing ticket; (ii) that the subject doesn't know that she is holding a losing ticket; and (iii) that the case is not a Gettier-case. That construal of lottery cases entails that *any* account of knowledge as justified, un-gettierized true belief is incorrect. So either some of the standard assumptions about lottery cases have to be given up or any account of knowledge as justified, un-gettierized true belief is wrong. In either case, reliabilism about justification is in the clear.

Finally, a problem raised by lottery considerations that is, *prima facie* at least, a serious problem for reliabilism has to do with *comparative* assessments of justification. Jonathan Adler has recently argued against reliabilism on the basis of this case:[14]

> A company that manufactures widgets knows that exactly one out of every thousand of their products suffers a singular defect as a by-product of (an ineliminable imperfection in) their excellent—and much better than average—manufacturing system. … Some managers would like to reduce the percentage of defects. The plan is to introduce a special detector, well designed to read 'OK' just in case the widget is not defective. …
>
> Smith and Jones, who both know of the one in a thousand defects, are each given a detector. …
>
> As each widget comes to his station, Smith momentarily glances away from the video game he is playing to stamp it 'OK', expressive of his corresponding (degree of) belief, while wholly ignoring the detector. As Smith knows, out of each batch of 1000, he is guaranteed to be correct 999 times by this method (and so better than by use of the detector, as explained next).
>
> … Knowing of her manager's well founded confidence in the detector, Jones applies it to each widget carefully and skillfully, and assigns it 'OK' (mostly) or 'Defect' (rarely) according to her determination. Given the complexity of operating the detector and normal human limits, the probability of an error in any evaluation is .003, though Jones is a first-rate technician.[15]

Adler claims that three things are true about his case. First, provided that she is not mistaken, Jones *knows* that the widgets that she stamps 'OK' are not defective, but Smith doesn't have the corresponding knowledge. Second, even though Smith can be justified in assigning a very high credence to the proposition that a given widget is not defective, he is not justified in '*all out* accepting it as true.'[16] Third, Adler thinks that all of this is true despite the fact that Smith uses a method that is more reliable than Jones's.

So Adler holds the fashionable view that a subject in a standard lottery case doesn't know that she is holding a losing ticket together with the unfashionable view that the same subject is not justified in thinking that she holds a losing ticket.[17] I will later argue (in section 5.3) that reliabilists have resources available to hold that Smith indeed isn't justified, if they wish to do so. But the claim that I am interested in now is the *comparative* claim that Jones is more justified than Smith when they both believe of a certain widget that it is OK.[18] There are two questions that we must ask here. First, is it really the case that Jones is more justified than Smith? Second, does reliabilism entail that Jones is not more justified than Smith? Regarding the first question, there are two ways of understanding Adler's case, and in one what way it is indeed true that Jones is more justified than Smith, whereas in the second it isn't. I'll say more about those two ways of understanding the case in the next section, when I examine the consequences that a probabilistic understanding of reliability has. But let me first address, in the remainder of this section, the issue of what are the consequences that a truth-ratio understanding of reliability has for the case.

For a variety of reasons, the answer to that question isn't straightforward. Notice, first, that reliabilism as I have defined it characterizes an all-or-nothing notion of justification: if the process that created the belief is sufficiently reliable, then the belief is justified, otherwise it isn't. It is natural, however, to think that it is in the spirit of reliabilism to claim that the more reliable the process, the more justified the belief is. Let us then construe reliabilism as incorporating this natural understanding of a degree-theoretic notion of justification.[19] Our question now is, then: who uses the more reliable process, Smith or Jones?

The answer to *that* question isn't straightforward either. In order to see why this is so, we must first determine what is the evidence that Jones and Smith use to arrive at their respective beliefs. Notice that, as Adler describes the case, Smith acquires his beliefs inferentially. Smith's evidence is constituted by his (justified) belief that any given widget belongs to a class of widgets of which exactly 1 in 1000 are defective. In Jones' case, her evidence is constituted by her justified belief (or perhaps her visual experience) that the detector reads 'OK'.[20]

According to both of the definitions of reliability in terms of truth-ratios, Smith's procedure has a 0.999 degree of reliability (Smith is *guaranteed* to be right *exactly* 999 out of 1000 cases). The consequence that Smith is justified in believing that a specific widget is OK can be avoided by setting r to a number higher than 0.999—but it is easy to see that, no matter how high *r* is set, we can always create a case where the procedure used by a counterpart of Smith is reliable to at least that degree.

What about Jones? What is the truth-ratio of the process of believing that a widget is OK based on a justified belief that the detector says that it OK? The answer is not as easy in this case as it was with Smith. All that Adler tells us about the detector is that the ‘probability of error’ in using it is 0.003. Adler clarifies that by probability of error he means both the probability that the procedure indicates that a widget is OK when it is defective as well as the probability that the procedure indicates that a widget is defective when it is OK. Notice that these probabilities don't *guarantee* anything about the actual truth-ratio of Jones’s beliefs. It is compatible with this information that Jones correctly classifies *all* of the widgets, as well as that she correctly classifies *none* of them—although this latter possibility is, of course, much less probable.

Does **reliability as high counterfactual truth-ratio** deliver a result for Jones? We might be tempted to think that it does, because we might be tempted to think that to say that a certain procedure has a 0.003 probability of error is to say that applications of that procedure will issue in a false belief in exactly 3 out of 1000 possible worlds. However, we must remember that, according to **reliability as high counterfactual truth-ratio**, not all possible worlds where the procedure is applied should be counted towards establishing the reliability of the procedure. Even if there is a direct translation from probabilities to truth in possible worlds, those 3 in 1000 worlds where the procedure results in false beliefs might count as too weird (too improbable!) to be ‘close’ to the actual world.

So, even though it is true that defining reliability in terms of truth-ratios gives us the result that Smith is justified, it doesn't give us the *comparative* result that Smith is more justified than Jones—simply because it is not clear what the truth-ratio of Jones’s procedure is. This is a serious problem for any understanding of reliability in terms of truth-ratios, which must be added to the conceptual problems for that understanding presented in section *truth-ratio*.

5.2 Lotteries and Reliability as High Conditional Probability

What are the consequences of *reliability as high conditional probability* with respect to Adler's case? It might already be obvious that the result, in the case of Smith, is 0.999, but it is not so obvious what the number is for Jones. The source of this obscurity is an ambiguity in the interpretation of the case. But, however the case is interpreted, *reliability as high conditional probability* is in the clear.

Let us first ask how we can compute the conditional probabilities in question (that is, the conditional probability that a given widget is OK given Smith's evidence and the conditional probability that it is OK given Jones's evidence). In the usual Kolmogorov axiomatization of the probability calculus, conditional probability is defined in terms of unconditional probabilities, as follows:

$$\Pr(A \mid B) =_{df.} \frac{\Pr(A \wedge B)}{\Pr(B)};$$

provided that Pr(B) > 0 (and otherwise the conditional probability is undefined). Bayes's theorem, in turn, states a main consequence of such definition:

$$\Pr(A \mid B) = \frac{\Pr(A)\Pr(B \mid A)}{\Pr(B)};$$

from where it is easy to see that the conditional probability of *A* given *B* is a function of the *unconditional* probabilities of *A* and *B* and the inverse conditional probability of *B* given *A* (what in the Bayesian jargon is called the 'likelihood' of *A* given *B*). Moreover, if we know the likelihood of the hypothesis on the evidence as well as the likelihood of the *negation* of the hypothesis given the evidence, we do not even need to know the unconditional probability of *B*, if we re-write the denominator in Bayes's theorem according to the 'total probability theorem':

$$\Pr(A \mid B) = \frac{\Pr(A)\Pr(B \mid A)}{\Pr(A)\Pr(B \mid A) + \Pr(\neg A)\Pr(B \mid \neg A)}$$

In Smith's case, we have all the information we need to calculate the conditional probabilities. We know that the probability that a particular widget belongs to a class of widgets of which exactly 1 in 1000 is defective is the same whether that widget is OK or not, something very close to 1. So, where 'W+' abbreviates the proposition that widget n is OK and 'S' abbreviates the proposition that widget n belongs to a class of widgets of which exactly 1 in 1000 is defective:

$$\Pr(W+\mid S)\approx\frac{0.999}{0.999+0.001}=0.999$$

We now need to know what is the probability that a widget is OK given that Jones's detector says that it is OK. There are here two ways of understanding Adler's case. According to the first one, Jones *is* more justified than Smith, but his procedure is also more reliable. According to the second one, Jones's procedure *isn't* more reliable that Smith's, but then Jones isn't more justified than Smith either. In either case, there is no problem for reliabilism.

The two ways of understanding Adler's case arise from an unclarity in Adler's presentation. What does he mean when he says that the 'error rate' of the detector is 0.003? The most natural way of understanding that claim is as giving us the probability that the detector says that widget is OK (defective) when it is actually defective (OK). That is, under this first understanding of the case, we have it that (letting 'D+' be the proposition that the detector says that widget n is OK, 'D-' the proposition that the detector says that widget n is defective, 'W-' the proposition that widget n is defective and, as before, 'W+' the proposition that widget n is OK):

$$\Pr(D+\mid W-)=\Pr(D-\mid W+)=0.003$$

Under this understanding of the case, it is true that we would naturally judge that Jones is more justified than Smith in his beliefs regarding widgets. But **reliability as high conditional probability** does have this consequence:

$$\Pr(W+\mid D+)=\frac{0.999\times 0.997}{0.999\times 0.997+0.001\times 0.003}=0.999997$$

So, no matter what *r* is, we do get the comparative result that Jones's procedure is more reliable than Smith's, and so Jones does count as justified if Smith does.[21] Whether either of them is justified depends, of course, on the choice of *r*, but it seems plausible to suppose that Jones at least will count as justified on any reasonable choice of *r*. Notice, too, that **reliability as high conditional probability** also gives the right result for a proposition that Adler doesn't consider: would Jones be justified in believing that a certain widget is *not* OK, based on the fact that the detector says so? Well, given that the error rate of the detector is 0.003 and that only 1 in 1000 widgets are not OK, on average, 3 out of 4 widgets that the detector indicates as being defective are actually OK. Given this, Jones shouldn't trust the detector when it says that a widget is defective—that is, if she does believe that a given widget is defective based on the detector's verdict, her belief is not justified.[22] And Pr(W-|D-) is quite low (roughly, 0.25), which means that, as desired, Jones is not justified in believing of a certain widget that it is not OK.

Let us turn now to the second understanding of Adler's case.[23] Under this interpretation, the 0.997 figure is not the be understood as giving us a likelihood, but rather as a *stipulation* of the conditional probability that a certain widget is OK given that the detector says that it is OK. This interpretation is very hard to square with what Adler actually *says*, but let us nevertheless consider it on its own merits. The suggestion, then, is that the description of the case stipulates that

$$\Pr(W+ \mid D+) = 0.997$$

If this is the correct understanding of the case, then there is no denying that Smith's procedure is more reliable than Jones's—at least under the definition of reliability under consideration. But probabilities cannot be stipulated in an unconstrained manner. As revealed by Bayes's theorem, conditional probabilities constrain (and are constrained by) prior probabilities (the probability that a widget is OK) and likelihoods. In Adler's case, the prior probability is fixed at 0.999. So, stipulating that the conditional probability that a widget is OK given that the detector says that it is OK is 0.997 constrains the conditional probabilities that we can assign to the detector's saying that a widget is OK given that it is OK, as well as the conditional probability that the detector says that a widget is OK given that it is not OK. There is no unique number that we *have* to assign to those probabilities, but all the possible assignments are discouraging. For instance, the following is one such assignment:

$$\Pr(D+ \mid W+) = 0.35$$

$$\Pr(D+ \mid W-) = 0.999$$

Once aware that the detector has such a high error rate, few would still judge that Jones is more justified than Smith in thinking of a particular widget that it is OK. One way to bring out the fact that we wouldn't now judge that Jones is more justified than Smith is by considering what Jones's judgments will be with respect to particular batches of 1,000 widgets.[24] For any given batch, 999 of 1,000 widgets will be OK. Of those, the detector will correctly classify as OK approximately 350—that is, it will misclassify approximately 650. What's even more disturbing, out of 1,000 widgets that are defective, the detector will misclassify as OK 999. That is, trusting the detector is *worse* than tossing a coin and believing that a widget is OK if and only if the coin comes up heads! Reliabilism does indeed entail that, under this interpretation, Jones is worse justified than Smith, but that, it seems, is as it should be.

Stipulating that Pr(W+|D+) = 0.997 has the consequence that the likelihood must be very low because the prior probability is 0.999. Couldn't we change the example so that the detector still has a reliability of 0.997 but the prior is much lower, so that the likelihood is also high? We could, of course. For instance, if we stipulate that fully half of the widgets are defective, then the reliability of the detector can be 0.997 while Pr(D+|W+) = 0.997, whereas Pr(D+|W-) = 0.003. But now, of course, Smith will be wrong half of the time, and so will come out as being less justified than Jones according to reliabilism.

Why do we tend to judge, with Adler, that Jones is better justified than Smith? It cannot be just because Jones uses a detector—the detector might well be disastrous (as with the second interpretation of the case). The reason for our judgment, I submit, is that we interpret the 0.997 figure as giving us the likelihood that the detector will say that a widget is OK when it is, and thus we correctly assume that the conditional probability is higher for Jones than it is for Smith. If we are explicitly told that 0.997 *is* the conditional probability for Jones, then (as soon as we realize what that entails about the detector) we no longer think that Jones is better justified than Smith. That is, there is no interpretation of the case where it is true that Jones is better justified than Smith but **reliability as high conditional probability** doesn't have this conclusion.

Now, we could, if we want, compare the justification that Smith has in the original case with Jones's justification in the modified case, where half of the widgets are defective.[25] Smith's procedure will still have a reliability of 0.999 whereas Jones's procedure will have a reliability of 0.997, and so, according to reliabilism, Smith is more justified than Jones. Again, this is as it should be. Of course, if Smith were to use his procedure at Jones's factory, then he *would* be less justified than Jones, and if Jones were to use his procedure at Smith's factory, then he *would* be more justified than Smith—and reliabilism does have that consequence. But this shouldn't make us give up the intuitive claim that Smith's procedure *in Smith's factory* results in beliefs that are more justified than Jones's procedure *in Jones's factory*.

What this last case does help to bring out, however, is the fact that it is not the case that Smith will count as *un*justified on any reasonable choice of *r*—as was the case with the definitions of reliability in terms of truth-ratios, it is possible to construct a lottery case with enough tickets to guarantee that Smith's procedure is reliable to any degree that we wish (in the sense of having a conditional probability at least that high). If one thinks that Smith may well be justified, and particularly if one thinks that as the numbers get higher the more justified Smith is, then the definition of reliability as high conditional probability is the correct one.

5.3 Getting Smith to Come Out as Unjustified

But what if one follows Adler in thinking that Smith is not even justified in believing that a given widget is OK?[26] In that case, is reliabilism shown to be false? No: There are plausible ways of defining reliability in terms of probability that do have the consequence that Smith is unjustifed.

Defining reliability in terms of conditional probability is closely related to Carnap's notion of 'confirmation as firmness.'[27] There is another quantity that Carnap also defined, 'confirmation as increase in firmness,' and which can serve as the basis for a different definition of reliability:

> **Reliability as increase in probability**: a type of the form *believing that p based on e* is reliable if and only if $\Pr(p \mid e) - \Pr(p) > r$.[28]

Applying **reliability as increase in probability** to Adler's case gives us the following results:

$$\Pr(W+ \mid S) - \Pr(W+) = 0.999 - 0.999 = 0$$
$$\Pr(W+ \mid D+) - \Pr(W+) = 0.9999997 - 0.999 = 0.000997$$

This means that, independently of the choice of r, Smith is not justified in his belief—the result that we were looking for. Whether Jones is justified or not depends, again, on the choice of r. And, even though 0.000997 seems too low to make Jones come out as justified, it should be noticed that r need not take the same value for all cases, and it can be a function, for instance, of $\Pr(p \mid e)$—it wouldn't be far-fetched, for example, to suggest that r must be inversely proportional to $\Pr(p \mid e)$ (which is itself directly proportional to $\Pr(p)$). For instance, if we set $r = 1 - \Pr(p \mid e)$, r will be 0.001 for Smith and 0.000003 for Jones. So, in so far as Adler's case goes, **reliability as increase in probability** gives the right results.[29]

One feature of Adler's case is that it is constructed so that it is clear what is the unconditional probability that a given widget is OK (or defective). But, as a rule, this will not be so. The problem is that, in many cases, such unconditional probabilities will not be available. Take a paradigmatic case of belief formation: I take a glance outside my study window and, as a result of having certain characteristic experiences, come to believe that there are flowers in the yard. What is the unconditional probability that there are flowers in the yard? And the unconditional probability that I will have those experiences? The problem is not that these questions are *hard* to answer; it is, rather, that they have no answer.[30] We can, of course, make sense of different questions: what is the probability that there are flowers in the yard, given that it is May? What is the probability that I will have certain experiences, given that I look in a certain direction? But, again, if the usual definition of conditional probability is correct, then the answer to *these* questions depends on the answer to the previous ones.[31] **Reliability as increase in probability** may be correct as a definition of reliability when it makes sense to think that there are unconditional probabilities to be had, as is the case in Adler's example. But cases of that sort will be the exception, not the rule.

We are not forced to accept the usual definition of conditional probability in terms of unconditional probabilities. There are other axiomatizations available that take the notion of conditional probability as *primitive*. When it does make sense to think of the unconditional probabilities, then Bayes' theorem can be seen as a *constraint* on conditional probabilities, but when it doesn't make sense we can still meaningfully talk of conditional probabilities.[32] Nevertheless, despite the fact that talking of conditional probabilities doesn't commit us to unconditional probabilities, **reliability as increase in probability** *itself* traffics in unconditional probabilities. That is reason enough to look elsewhere for a definition of reliability to use when unconditional probabilities are not available.[33]

Reliability as increase in probability is one of the many measures of *confirmational relevance* that have been proposed in the literature on Bayesianism.[34] The interesting question for us now, then, is whether there is a measure that doesn't itself involve reference to unconditional probabilities. And the answer is, fortunately, 'Yes.' In particular, what is called the 'likelihood ratio' measure provides us with the basis for a definition of reliability that doesn't rely on unconditional probabilities:

> **Reliability as high likelihood ratio**: a type of the form *believing that p based on e* is reliable if and only if $\frac{\Pr(e \mid p)}{\Pr(e \mid \neg p)} > 1 + r$

for some suitable r. When the proposition in question just doesn't have an unconditional probability, the right definition of reliability to use is not **reliability as increase in probability** but **reliability as high likelihood ratio**.

Now, what does **reliability as high likelihood ratio** say with respect to Adler's case? For the proposition that a certain widget is OK, it gives the results that we are looking for in this section:

$$\frac{\Pr(S \mid W+)}{\Pr(S \mid W-)} \approx \frac{1}{1} = 1$$

$$\frac{\Pr(D+ \mid W+)}{\Pr(D+ \mid W-)} = \frac{0.997}{0.003} = 332.333$$

No matter how small we set r, Smith will not count as justified under this definition. Intuitively, Smith's procedure has a likelihood ratio of 1 because it is *insensitive* to whether any particular widget is OK or not. Jones, by contrast, will count as justified under any plausible value for r. The high likelihood ratio of Jones's procedure reflects the fact that *it is* highly sensitive to whether a particular widget is OK or not.[35]

6. Conclusion

If we think that subjects like Smith can be justified, then **reliability as high conditional probability** results in a reliabilist theory that is correct as far as cases of the sort examined in this paper go. (We must remember, though, that to apply the resulting theory to cases where there are no unconditional probabilities to be had we should not think of conditional probability as defined in terms of unconditional probability.) If, on the other hand, we think that subjects like Smith cannot be justified, then a combination of **reliability as increase in probability** and **reliability as high likelihood ratio** delivers the right results for these kinds of cases. It remains to be seen whether a reliabilist theory armed with these definitions is free from other kinds of counterexamples, or whether it has any other flaws.[36] But it does show that the common belief that reliabilism has *special* problems with lottery cases is simply false.[37]

Philosophy Department

University of Wisconsin – Madison

Notes

---

[1] Goldman (1979), the *locus classicus* for reliabilism, was aware of the problem, and it has been forcefully pressed by Conee and Feldman (1998). Reliabilism is also subject to apparent counterexamples, in the guise of cases that challenge both the sufficiency and the necessity of reliability for justification. See note 36.

[2] For instance, the token process that produced my belief that there are flowers in the yard belongs to these types, among many others: visual process, hasty scanning of my visual field, perceptual process, etc.

[3] See Comesaña (2006) for details and arguments.

[4] See Conee and Feldman (1985).

[5] I develop and defend this combination of reliabilism and evidentialism (which I call 'Evidentialist Reliabilism') in Comesaña (forthcoming).

[6] Notice two things about clause (ii) in this definition. First, the notion of justification itself appears in this clause. However, this is just a convenient way of representing the difference between inferential and non-inferential beliefs, and doesn't make the definition objectionably circular. Second, when the belief in question is inferentially acquired (when it is based, in part, on other justified beliefs of the subject), the connection that has to be reliable is that between the propositional content of the justified beliefs that the subject uses as evidence and the target belief in question. Thus, if I base my belief that the streets are wet on my justified beliefs that it is raining and that if it is raining then the streets are wet, then my belief is justified (according to the version of reliabilism that we are considering) if and only if the type *my believing that the streets are wet based on the facts that it is raining and that if it is raining then the streets are wet* is reliable. This has the equivalent effect of appealing to *conditional* reliability, as in Goldman (1979).

[7] See Baumann (1998). The project of basing justification on objective probabilistic connections is obviously related to the so-called 'logical' interpretation of probability championed by Keynes (1921) and Carnap (1950), but it is important to realize that the project need not take the form of presenting a probability function over sentences of a language that takes into account only *syntactic* features of the sentences, à la Carnap. In fact, the objective probability in question need not even be representable by a probability function, or even a set of probability functions. For a defense of this project, see Part Two of Stove (1986).

[8] Alston (1988), pp. 217-18 of the version in Bernecker and Dretske (2000), emphasis in the original.

[9] See Kyburg (1961) for the classic treatment of this problem. For a recent discussion that highlights the resilience of the problem, see Douven and Williamson (2006).

[10] For a detailed examination of the closure problem as it applies to lotteries, see Hawthorne (2004).

[11] Or, as Hawthorne (2004) notes, on the fact that we won't be able to afford the African Safari.

[12] Even Hawthorne (2004), who claims that reliabilism *as a theory of knowledge* does very badly at predicting our intuitions in lottery cases (p. 9), thinks that 'lottery propositions that we claim not to know may both be true and well justified' (p. 8).

[13] In the canonical Gettier cases, the bad luck consists in having a justified *false* belief, whereas the good luck consists in nevertheless inferring a *true* proposition from that false belief.

[14] Earlier, Stewart Cohen had argued on similar grounds that all fallibilist theories of knowledge run into trouble—see Cohen (1988).

[15] Adler (2005), pp. 446-7.

[16] Adler (2005), p. 448.

[17] One potentially confusing factor in Adler's case is that what Smith does has *bad consequences* for the factory he works at, for he ships *all* of the defective widgets (whereas Jones probably doesn't ship *any* defective widget). Would Adler's intuitions change if only one in a thousand widgets is OK, and Smith believes of each of them that it is defective? And even in the original case, would Adler hold on to his intuitions as the numbers get higher: what if only one in a billion widgets is defective?

[18] It is important to note that, as explained below, Jones is *not* justified in believing of a certain widget that it is defective on the basis of her detector's saying so.

[19] Notice, however, that adhering to that characterization of a degree-theoretic notion of justification doesn't commit the reliabilist to the claim the difference in reliability corresponds to an analogous difference in justification.

[20] Together, perhaps, with her justified belief that the detector is reliable. But what we are trying to find out is precisely what it is for a procedure to be reliable.

[21] In favor of this interpretation of the case, notice that Adler himself recognizes that the probability that a certain widget is OK given that the detector says it is OK is 0.999997 (Adler (2005), p. 452), but he doesn't seem to realize that this means that Jones is more reliable than Smith according to a very natural way of measuring reliability.

[22] This is not to say that she shouldn't throw away any widget that the detector indicates as defective—she should, if her goal is to get rid of as many defective widgets as she can.

[23] Michael Levin suggested this interpretation in comments to a previous version of the paper.

[24] What follows assumes that the frequencies match up with the probabilities in this case, but no general theory of probabilities as frequencies is assumed.

[25] This suggestion is due to Michael Levin.

[26] Why would one think that? One possible motivation is mistrust of so-called *purely statistical* evidence. I think that the mistrust is misplaced, but I won't argue the point here.

[27] Carnap (1962), new preface.

[28] Sturgeon (2000), pp. 96-8 rejects both **reliability as high conditional probability** and **reliability as increase in probability**. His objection to **reliability as high conditional probability** is that it doesn't ensure that the proposition believed will be made more probable than it was before, and his objection to **reliability as increase in probability** is that it need not made the proposition very probable at all. The first objection can be answered by noticing that, in our definition, the belief is justified when it is *based* on evidence conditional on which it has a high probability. This is as it should be, even if there is some other evidence which the subject doesn't avail himself of on which the proposition has even higher probability—and if the subject does avail himself of that additional evidence, then of course he will also be justified according to **reliability as high conditional probability**. The second objection is taken care of by the suggestion in the text that *r* need not be constant, and may instead be a function of the prior probability of the proposition in question.

[29] Also, note that setting *r* again as $1-\Pr(p \mid e)$ (in this case, = 0.75), **reliability as increase in probability** gives the correct result that Jones would not be justified in believing that a widget is not OK on the basis of the detector's saying so:

$$\Pr(W- \mid D-) - \Pr(W-) = 0.25 - 0.001 = 0.249$$

[30] See Hájek (2007), Hájek (2003), Sober (2002) and Sober (2008).

[31] Actually, the unconditional probability of the evidence is a function of the conditional probability of the evidence on each of the competing hypotheses, if they form a mutually exclusive and jointly exhaustive set, as the 'total probability' theorem shows. But this is

no comfort to the enemy of unconditional probabilities, because the total probability theorem establishes that the unconditional probability of the evidence is a function of the conditional probabilities of the evidence on the hypotheses only together with the *unconditional* probabilities of these hypotheses. Which means that, in our case, we would still have to hunt for the elusive unconditional probability that there are flowers in the yard.

[32] Hájek (2003).

[33] But notice that the availability of a conception of probability in which conditional probability is primitive *is* a solution to the problem of the unavailability of some unconditional probabilities if we are happy with **reliability as high conditional probability**.

[34] See Fitelson (1999).

[35] **Reliability as high likelihood ratio** also gives the result 332.333 for Jones's belief that a certain widget is not OK based on the detector's saying so. This is the wrong result in this case, but that is because we are applying **reliability as high likelihood ratio** when we should be applying **reliability as increase in probability** or **reliability as high conditional probability** (because, again, in Adler's cases we do have available the unconditional probability that a widget will be OK). Because the probability that a widget is defective is so low, most of the widgets that the detector says are defective are actually OK. If this prior probability were simply not to be had, then we would judge that Jones should trust the detector whatever its verdict.

[36] There are two main sources of counterexamples to reliabilism. First, what has come to be known as 'the new evil demon problem' (see Cohen (1984) and Sosa (1991)), which challenges the necessity of reliability for justification by noting that the beliefs of a victim of an evil demon are as justified as ours and yet (the objector claims) are not reliably formed. I argue for a version of reliabilism that can answer this challenge in Comesaña (2002). Second, there are cases such as those presented in BonJour(1980) that

challenge the sufficiency of reliability for justification: a very reliable clairvoyant with no evidence of his clairvoyant powers is not justified in his beliefs, despite the fact that they are indeed very reliably produced. The combination of reliabilism with evidentialism presented here (and more fully developed in Comesaña (forthcoming)) has the resources necessary to deal with this kind of counterexample, for BonJour's subject forms his belief on the basis of no evidence at all.

[37] Many thanks to Jonathan Adler, Michael Levin (my commentator on a previous draft of this paper that was presented at the 2006 Pacific APA), Elliott Sober, Carolina Sartorio, Peter Vranas and an anonymous referee for the *Pacific Philosophical Quarterly* for very valuable comments on previous versions of this paper.

References

Adler, J. (2005) "Reliabilist Justification (or Knowledge) as a Good Truth-Ratio," *Pacific Philosophical Quarterly* 86:4, 445-458.

Alston, W. (1988) "An Internalist Externalism," *Synthese* 74, 265-283.

Baumann, P. (1998) "Can Reliabilists Believe in Subjective Probability?," *The Philosophical Quarterly* 48:191, 199-200.

Bernecker, S., and F. Dretske, eds. (2000) *Knowledge*, Oxford University Press.

BonJour, L. (1980) "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy* V, 53-73.

Carnap, R. (1950) *Logical Foundations of Probability*, University of Chicago Press.

Carnap, R. (1962) *Logical Foundations of Probability*, University of Chicago Press, 2nd. Edition.

Cohen, S. (1984) "Justification and Truth," *Philosophical Studies* 46, 279-295.

Cohen, S. (1988) "How To Be A Fallibilist," *Philosophical Perspectives* 2, 91-123.

Conee, E. and R. Feldman "Evidentialism," *Philosophical Studies* 48. Reprinted with an "Afterword" in Conee and Feldman (2004), pp. 83-107.

Conee, E., and R. Feldman (1998) "The Generality Problem for Reliabilism," *Philosophical Studies* 89, 1–29.

Conee, E., and R. Feldman (2004) *Evidentialism*, Oxford University Press, New York.

Comesaña, J. (2002) "The Diagonal and the Demon," *Philosophical Studies* 100, 249-266.

Comesaña, J. (2006) "A Well-Founded Solution to the Generality Problem," *Philosophical Studies* 129, 27-47.

Comesaña, J. (forthcoming) "Evidentialist Reliabilism," *Noûs*.

Douven, I., and T. Williamson (2006) "Generalizing the Lottery Paradox," *British Journal for the Philosophy of Science* 57:4, 755–779.

Fitelson, B. (1999) "The plurality of Bayesian measures of confirmation and the problem of measure sensitivity," *Philosophy of Science* 66, S362–S378.

Goldman, A. (1979) "What is Justified Belief?" in G. Pappas, ed., *Justification and Knowledge*, Reidel, Dordrecht, 1–23.

Hàjek, A. (2003) "Conditional Probability is the Very Guide of Life," in H. Kyburg and M. Thalos, eds., *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, Open Court.

Hàjek, A. (2007) "The reference class problem is your problem too," *Synthese* 156:3, 563-585.

Hawthorne, J. (2004) *Knowledge and Lotteries*, Oxford University Press.

Keynes, J. (1921) *A Treatise on Probability*, Macmillian.

Kyburg, H. (1961) *Probability and the Logic of Rational Belief*, Wesleyan University Press.

Sober, E. (2002) "Bayesianism: Its Scope and Limits," in R. Swinburne, ed., *Bayes' Theorem*, Proceedings of the British Academy Press.

Sober, E. (2008) *Evidence and Evolution*, Cambridge University Press.

Sosa, E. (1991) *Knowledge in Perspective*, Cambridge University Press, New York.

Stove, D. C. (1986) *The Rationality of Induction*, Oxford University Press.

Sturgeon, S. (2000) *Matters of Mind*, Routledge.