This is a preprint of an article whose final and definitive form will be published in the *Australasian Journal of Philosophy*. The *Australasian Journal of Philosophy* is available online at: http://www.tandf.co.uk/journals/.

THE BRIER RULE IS NOT A GOOD MEASURE OF EPISTEMIC UTILITY (AND OTHER USEFUL FACTS ABOUT EPISTEMIC BETTERNESS)

Don Fallis and Peter J. Lewis

Measures of epistemic utility are used by formal epistemologists to make determinations of epistemic betterness among cognitive states. The Brier rule is the most popular choice (by far) among formal epistemologists for such a measure. In this paper, however, we show that the Brier rule is sometimes seriously wrong about whether one cognitive state is epistemically better than another. In particular, there are cases where an agent gets evidence that definitively eliminates a false hypothesis (and the probabilities assigned to the other hypotheses stay in the same ratios), but the Brier rule says that things have gotten epistemically worse. Along the way to this 'elimination experiment' counter-example to the Brier rule as a measure of epistemic utility, we identify several useful monotonicity principles for epistemic betterness. We also reply to several potential objections to this counter-example.

Keywords: Bayesianism, Brier score, epistemic utility, formal epistemology, law of likelihoods, proper scoring rules.

1. Introduction

Epistemology is a normative project. In particular, many epistemologists want to identify cognitive processes and social practices that lead to epistemically good cognitive states, such as true belief, knowledge, and understanding. But in order to do this, they have to be able to say when one cognitive state is epistemically better than another (see Goldman [1999]; Fallis and Whitcomb [2009]).

Admittedly, not all epistemologists endorse this sort of epistemic consequentialism (see, e.g., Kelly [2003]). This paper, however, is aimed at those epistemologists who do. Such epistemologists often use *measures of epistemic utility* to make these determinations of epistemic betterness among cognitive states. The *Brier rule* is the most popular choice (by far) among epistemologists for such a measure (see Kierland and Monton [2005: 385-387]; Greaves and Wallace [2006: 627-28]; Joyce [2009: 290-93]; Leitgeb and Pettigrew [2010a: 219-20]). In this paper though, we argue that the Brier rule is not a good measure of epistemic utility. We show that it is sometimes seriously wrong about whether one cognitive state is epistemically better than another. Along the way to this result, we identify several useful monotonicity principles for epistemic betterness.

2. Epistemic betterness

Although there are several other epistemic values (e.g. justification, understanding), epistemologists typically focus on the degree to which an agent's cognitive state gets at the *truth* (see Goldman [1999: 87-94]; Joyce [2009]; Leitgeb and Pettigrew [2010a]).[1] So, for instance, a true belief on some topic is epistemically better than a false belief, and suspending judgment falls somewhere between the two (see Goldman [1999: 89]; Fallis [2007: 222]).

For many purposes in epistemology, this simple epistemic ordering of categorical belief states is sufficient. However, many epistemologists these days represent an agent's cognitive

---

[1] Don Fallis and Dennis Whitcomb [2009] consider how epistemic values beyond truth might fit into the framework of epistemic consequentialism.

state in terms of her *credences* over a set of mutually exclusive and jointly exhaustive hypotheses (i.e. over a partition) (see Goldman [1999: 90]; Joyce [2009]; Leitgeb and Pettigrew [2010a]).[2] In other words, an agent's cognitive state is taken to be **r** = (r_1, r_2, ..., r_n), where r_i is the probability that she assigns to hypothesis h_i being true. For instance, suppose that there are three suspects (Tom, Dick, Harry) in a murder investigation. Further, suppose that the detective (Sam) currently thinks that each of the three suspects is equally likely to be guilty. In that case, **r** = (1/3, 1/3, 1/3) captures Sam's cognitive state.

Epistemologists need a way to rank credences in terms of how close they get to the truth. When there are just two hypotheses, there is a fairly straightforward epistemic ordering that seems to do the job correctly. Namely, if h_i is true, then **r** is epistemically better than **s** if and only if $r_i > s_i$. We might call this the *linear ordering* of credences. However, things start to get tricky when there are three or more hypotheses.

Even with three or more hypotheses, there are still many comparisons that are clear-cut. For instance, if h_1 is true, then **r** = (1/2, 1/4, 1/4) is epistemically better than **s** = (1/3, 1/3, 1/3). **r** assigns a higher probability to the true hypothesis than **s** does, and it assigns a lower probability to both of the false hypotheses. Similarly, if h_1 is true, then **r** = (3/4, 1/4, 0) is epistemically better than **s** = (1/2, 1/4, 1/4). In addition, if h_i is true, $r_i = 1$, and $r_j = 0$ for all $j \neq i$, then **r** is epistemically better than any other coherent credences.

But there are other comparisons that are more difficult. For instance, if h_1 is true, then is **r** = (1/2, 1/2, 0) or **s** = (1/3, 1/3, 1/3) epistemically better? **r** assigns a higher probability to the true hypothesis than **s** does, but it also assigns a higher probability to one of the false hypotheses

[2] For the sake of simplicity, we set aside the possibility that the different hypotheses might have different degrees of *informativeness* in this paper. We also set aside the possibility that the different *false* hypotheses might have different degrees of *verisimilitude*. See Graham Oddie [2014] for a discussion of formal models of verisimilitude or 'truthlikeness.'

(h_2). It is not immediately clear whether the epistemic advantages of **r** outweigh its apparent epistemic disadvantage.

3. Measures of epistemic utility

Coming up with a correct epistemic ordering of cognitive states is tricky enough. But epistemologists often want more. They want a *measure of epistemic utility* (see Maher [1990]; Oddie [1997]; Fallis [2007]; Joyce [2009]; Leitgeb and Pettigrew [2010a]). This is a function that yields a cardinal value that says how close an agent's credences are to the truth. This value can be plugged into a decision matrix, which can then be used to pick actions that will maximize expected epistemic utility.[3] For instance, Sam might use such a decision matrix to determine whether it would be epistemically beneficial (in expectation) to check for Harry's fingerprints on the murder weapon.

Measures of epistemic utility were originally used to explain why it is an epistemically good idea for scientists to perform experiments. Epistemologists (e.g., Maher [1990]; Oddie [1997]; Fallis [2007]) have argued that, if scientists conform to the various tenets of Bayesianism (e.g., coherence, conditionalization), then performing an experiment will always maximize expected epistemic utility. But epistemologists have also used measures of epistemic utility to provide justifications for the various tenets of Bayesianism themselves. For instance, some (e.g., Joyce [2009: 285-88]; Leitgeb and Pettigrew [2010b]) have argued that having *coherent* credences maximizes expected epistemic utility.[4] In addition, some (e.g., Oddie [1997: 541];

[3] Since we are focusing here on the degree to which an agent's cognitive state gets at the truth, we might speak instead about *measures of inaccuracy* and about *minimizing expected inaccuracy* as many formal epistemologists (e.g., Joyce [2009]; Leitgeb and Pettigrew [2010a]) do.
[4] Strictly speaking, Joyce's result is even stronger than this. For *measures of epistemic utility* that have certain independently desirable properties (such as *propriety* and *separability*), Joyce shows

Greaves and Wallace [2006]; Leitgeb and Pettigrew [2010b]) have argued that *conditionalizing* on new evidence maximizes expected epistemic utility.[5]

4. Proper scoring rules

In addition to providing an epistemic ordering of *basic* cognitive states (i.e. credences), a measure of epistemic utility needs to provide an epistemic ordering of *lotteries* over basic cognitive states (see Fallis [2007: 217-19]). For instance, while checking for Harry's fingerprints on the murder weapon will probably take Sam closer to the truth, she knows that there is a chance that doing so will take him further away (since fingerprint tests are not 100% reliable). Thus, Sam has to determine whether it is epistemically better to 'buy a ticket' for this epistemic lottery or to stick with her current cognitive state. In fact, since Sam is not yet certain which of the hypotheses is true, just sticking with her current credences is itself an epistemic lottery.

The need to provide a correct epistemic ordering of lotteries yields an important constraint on measures of epistemic utility. As most epistemologists agree, a measure of epistemic utility must be a *proper scoring rule* (PSR) (see Maher [1990]; Oddie [1997]; Greaves

---

that, for any credences **c** that are not coherent, there are coherent credences **c′** that have a higher epistemic utility than **c** regardless of which hypothesis happens to be true (see Pettigrew [2013: 900-02]). In other words, he uses *utility dominance* rather than *expected utility maximization* to vindicate *probabilism*.

[5] Goldman [1999: 115-23] argues that conditionalizing on new evidence maximizes *objectively* expected epistemic utility and not just subjectively expected epistemic utility. But as we discuss in the following section, the measure of epistemic utility that he uses to prove this result is not a proper scoring rule. So, it is not an appropriate measure of epistemic utility. Moreover, Goldman's result about conditionalization does not hold for any bounded proper scoring rule (see Fallis and Liddell [2002]).

and Wallace [2006]; Fallis [2007]; Joyce [2009]; Leitgeb and Pettigrew [2010a]). In other words, it must satisfy the *propriety* constraint that $\sum_k r_k u_k(\mathbf{r}) \geq \sum_k r_k u_k(\mathbf{s})$ for all **r** and **s**.[6]

If a measure of epistemic utility does not satisfy the propriety constraint, it will sometimes be the case that some other credences look epistemically better from the perspective an agent's current credences. Thus, such a measure of epistemic utility will sometimes say that it is epistemically beneficial for an agent to change her cognitive state in the absence of new evidence. For instance, an obvious candidate for a measure of epistemic utility is the *linear rule* (see Goldman [1999: 90]).

Linear rule: $u_i(\mathbf{r}) = r_i$, where $u_i(\mathbf{r})$ is the epistemic utility of **r** when h_i is true, and r_i is the probability assigned to h_i.

The linear rule yields the linear ordering of basic cognitive states discussed above. However, not being a PSR, the linear rule has the unfortunate feature that an agent can always maximize expected epistemic utility simply by assigning a probability of 1 to the hypothesis that she currently thinks is most likely to be true (see Maher [1990: 112-13]; Fallis [2007: 229]).

Fortunately, the propriety constraint is not overly demanding. There are infinitely many PSR. But researchers typically focus on three PSR: the Brier rule, the logarithmic rule, and the spherical rule (see Bickel [2007: 49]).[7]

---

[6] Note that this statement of propriety is not restricted to *coherent* credences. Also, some epistemologists (e.g., Oddie [1997: 539]; Joyce [2009: 276]) require that measures of epistemic utility be *strictly* proper scoring rules. That is, $\sum_k r_k u_k(\mathbf{r}) > \sum_k r_k u_k(\mathbf{s})$ for all **r** and **s**. All of the scoring rules under discussion in this paper are strictly proper.

[7] Note that the following statements of these rules presuppose that credences are over a partition. See Joyce [2009: 275] for statements of these rules for credences over a Boolean algebra. We discuss the issue of partitions versus Boolean algebras below. Also, the Brier rule is often given as a measure of inaccuracy (with the sign reversed) rather than as a measure of epistemic utility as it is here. It is sometimes referred to as the *quadratic rule*.

Brier rule: $u_i(\mathbf{r}) = 2r_i - \sum_k r_k^2$, where $u_i(\mathbf{r})$ is the epistemic utility of $\mathbf{r}$ when h_i is true, r_i is the probability assigned to h_i, and $\sum_k r_k^2$ is the sum of r_k^2 for all k such that $1 \leq k \leq n$ where n is the number of hypotheses in the partition.

Logarithmic rule: $u_i(\mathbf{r}) = \ln(r_i)$

Spherical rule: $u_i(\mathbf{r}) = r_i / \sqrt{\sum_k r_k^2}$

Which of these rules (if any) should we choose as our measure of epistemic utility?

5. Selecting a Measure

These three rules all have one property that would seem to be essential for a measure of epistemic utility. For all three rules, all other things being equal, $u_i(\mathbf{r})$ increases as r_i increases. In other words, the epistemic utility of an agent's credences goes up as the probability that she assigns to the true hypothesis goes up.

With the logarithmic rule, *only* the probability assigned to the true hypothesis matters in determining the epistemic utility of an agent's credences. So, it has the advantage of being a PSR that yields the linear ordering of basic cognitive states discussed above. However, as we explain below, there is also something to be said for taking into account how the probabilities assigned to *false hypotheses* are distributed.

The Brier rule and the spherical rule both include the term $\sum_k r_k^2$, which is the sum of the squares of the probabilities assigned to all of the hypotheses in the partition. This sum is smaller when the probabilities are more evenly distributed over more hypotheses. It is larger when the probabilities are concentrated on fewer hypotheses.[8] And both rules take such concentration to

[8] For coherent credences, this sum is minimized when a probability of $1/n$ is assigned to all of the hypotheses in the partition. It is maximized when a probability of 1 is assigned to one of the hypotheses.

be a negative when determining the epistemic utility of an agent's credences (the Brier rule by subtracting $\sum_k r_k^2$ and the spherical rule by dividing by the square root of $\sum_k r_k^2$).[9]

Of course, concentrating probability on the *true hypothesis* is clearly a good thing, epistemically speaking. Indeed, if a greater concentration is due solely to an increase in the probability assigned to the true hypothesis, it is a net epistemic benefit according to the Brier rule and the spherical rule. However, if a greater concentration is due solely to an increase in the probability assigned to a *false hypothesis*, it leads to a decrease in epistemic utility according to both rules.[10] This seems like a reasonable property for a measure of epistemic utility to have.[11] It is probably a big part of why the Brier rule is the PSR that is most often proposed as a measure of epistemic utility.

Even though their results usually only depend on its propriety, many epistemologists (e.g., Maher [1990: 113]; Oddie [1997: 538-39]; Greaves and Wallace [2006: 627-28]; Fallis [2007: 222]; Pettigrew [2013: 899-900]) give the Brier rule as their prime example of a measure of epistemic utility. Some epistemologists (e.g., Joyce [2009: 290-93]; Leitgeb and Pettigrew [2010a: 219-20]) go further and argue that the Brier rule has features that make it the best candidate for a measure of epistemic utility. For instance, Hannes Leitgeb and Richard Pettigrew claim that 'global inaccuracy of a global belief function *b* at a world *w* ought to be a strictly increasing function only of the Euclidean distance between the vector representation of *b* and the

---

[9] Since they handle this term a bit differently, the two rules do give a different weight to this epistemic cost.

[10] If the probability assigned to the true hypothesis *and* the probability assigned to a false hypothesis both increase, epistemic utility may increase or decrease depending on the details of the case.

[11] In 'A Strange Thing about the Brier Score,' Brian Knab and Miriam Schoenfield [2015] suggest that 'falsity distributions don't matter.' Lewis and Fallis [2014] argue that they do matter. But this debate is orthogonal to our concerns here. It is the *specific way* that the Brier rule handles falsity distributions that we object to in this paper.

vector representation of *w*.' (James Joyce's 'Homage to the Brier Score' [2009: 290] is another notable example here.) But in addition, at least a few epistemologists (e.g., Kierland and Monton [2005]; Leitgeb and Pettigrew [2010b]) use the Brier rule to derive their results about epistemic utility.

Only a few philosophers (e.g., Levinstein [2012]; Lewis and Fallis [2014]; Knab and Schoenfield [2015]) have criticized the Brier rule as a measure of epistemic utility. For instance, Ben Levinstein shows that it requires us to reject Jeffrey Conditionalization in favor of a much less attractive updating procedure. But the critique of the Brier rule that we offer below is more fundamental.[12]

In this paper, we show that the Brier rule does not even provide a correct epistemic ordering of basic cognitive states (even when we restrict our attention to *coherent* credences as we will do here). In particular, it incorrectly says that **s** = (1/4, 1/2, 1/4) is epistemically better than **r** = (1/3, 2/3, 0) when h_1 is true. (We explain below why this is a mistake.) Thus, we conclude that epistemologists who want a measure of epistemic utility should choose among the other available proper scoring rules.

6. Monotonicity principles

Before we get to the problem case for the Brier rule, let us briefly consider the clear-cut comparisons mentioned above. Suppose that h_1 is true, **r** = (1/2, 1/4, 1/4), and **s** = (1/3, 1/3, 1/3).

[12] Like other PSR, the Brier rule is typically used as a method for eliciting probability estimates (see Bickel [2007]). Our criticism though is just to the Brier rule as a measure of epistemic utility.

As noted above, **r** is clearly epistemically better than **s**. The Brier rule, the logarithmic rule, and the spherical rule all agree with this intuitive judgment.[13]

This comparison between **r** and **s** is actually a special case of an attractive *monotonicity principle*:

> **M1**. All other things being equal, if **r** assigns a higher probability to the true hypothesis than **s** does, then **r** is epistemically better than **s**.

Like all the monotonicity principles that we discuss in the paper, **M1** asks us to consider two coherent credences (**r** and **s**) that differ from each other in a specific way. It then tells us which of the two credences has a higher epistemic utility on the assumption that a particular hypothesis (from the partition of hypotheses that the agent is considering) is true. What 'all other things being equal' means in the context of coherent credences is that the probabilities assigned to the other hypotheses are all in the same ratios.[14] So, more formally, what **M1** says is that, for all i such that $1 \leq i \leq n$ (where n is the number of hypotheses in the partition), if $r_i > s_i$, and there is a real number α such that, for all $j \neq i$ such that $1 \leq j \leq n$, $\alpha r_j = s_j$, then $u_i(\mathbf{r}) > u_i(\mathbf{s})$. As noted above, the Brier rule, the logarithmic rule, and the spherical rule endorse this principle. Indeed, *all* PSR endorse **M1** (see Fallis [2007: 240-41]).

There are other clear-cut cases though that are not captured by **M1**. For instance, suppose that h_1 is true, **r** = (3/4, 1/4, 0), and **s** = (1/2, 1/4, 1/4). In this case, the probabilities assigned to the false hypotheses are not in the same ratios in **r** and **s**. Even so, **r** is clearly

[13] According to the Brier rule, $u_1(\mathbf{r}) = 0.625$ and $u_1(\mathbf{s}) = 0.333$. According to the logarithmic rule, $u_1(\mathbf{r}) = -0.301$ and $u_1(\mathbf{s}) = -0.477$. According to the spherical rule, $u_1(\mathbf{r}) = 0.817$ and $u_1(\mathbf{s}) = 0.577$.

[14] Since the probabilities have to sum to 1 for coherent credences, the other probabilities cannot all be exactly the same. That is, if r_i is different than s_i, then some of the other probabilities in **r** and **s** have to be different as well.

epistemically better than **s**. The Brier rule, the logarithmic rule, and the spherical rule all agree with this intuitive judgment.[15]

This comparison between **r** and **s** is actually a special case of an even stronger monotonicity principle:

M2. If **r** assigns a higher probability to the true hypothesis than **s** does, and **r** does not assign a higher probability to any false hypothesis, then **r** is epistemically better than **s**.

More formally, if $r_i > s_i$, and $r_j \leq s_j$ for all $j \neq i$, then $u_i(\mathbf{r}) > u_i(\mathbf{s})$. The Brier rule, the logarithmic rule, and the spherical rule all endorse this principle.[16]

7. Elimination experiments

We will now show that the more difficult comparison mentioned above is not actually that difficult. Suppose that h_1 is true, $\mathbf{r} = (1/2, 1/2, 0)$, and $\mathbf{s} = (1/3, 1/3, 1/3)$. The Brier rule, the logarithmic rule, and the spherical rule all agree that **r** is epistemically better than **s**.[17]

Moreover, there is a good reason to think that **r** *is* epistemically better than **s** when h_1 is true. Recall that Sam starts out thinking that each of the three suspects is equally likely to be guilty. Suppose that she then gets evidence that *definitively* eliminates Harry as a suspect and that she conditionalizes on this evidence. In that case, Sam's cognitive state goes from $\mathbf{s} = (1/3, 1/3, 1/3)$ to $\mathbf{r} = (1/2, 1/2, 0)$. It seems clear that, regardless of whether Tom or Dick is guilty, this

---

[15] According to the Brier rule, $u_1(\mathbf{r}) = 0.875$ and $u_1(\mathbf{s}) = 0.625$. According to the logarithmic rule, $u_1(\mathbf{r}) = -0.125$ and $u_1(\mathbf{s}) = -0.301$. According to the spherical rule, $u_1(\mathbf{r}) = 0.949$ and $u_1(\mathbf{s}) = 0.817$.

[16] This claim is trivial for the logarithmic rule. Proofs of this claim for the Brier rule and for the spherical rule are in the appendix. It may not be the case though that all PSR endorse this principle.

[17] According to the Brier rule, $u_1(\mathbf{r}) = 0.500$ and $u_1(\mathbf{s}) = 0.333$. According to the logarithmic rule, $u_1(\mathbf{r}) = -0.301$ and $u_1(\mathbf{s}) = -0.477$. According to the spherical rule, $u_1(\mathbf{r}) = 0.707$ and $u_1(\mathbf{s}) = 0.577$.

constitutes evidence-driven progress towards the truth (or at least away from falsity). Indeed, such 'elimination experiments' are a standard part of scientific practice (see Earman [1992: 163-85]). As John Earman [1992: 165] notes, 'even if we can never get down to a single hypothesis, *progress* occurs if we succeed in eliminating finite or infinite chunks of the probability space' (emphasis added).[18]

Admittedly, the probability that Sam assigns to one false hypothesis (viz. that Dick is guilty) does go up. This fact would lead some philosophers (e.g., Skyrms [2010: 80]; Smead [2014: 858]; Bruner [2015: 659-60]; Martínez [2015: 217-18]) to conclude that Sam has been misled.[19] However, there is really no epistemic downside here. Although the probability that Sam assigns to a false hypothesis goes up, the probability that she assigns to the true hypothesis goes up in the same ratio. In other words, and assuming that we measure degree of confirmation using *likelihoods* as many formal epistemologists (e.g., Hacking [1965: 70]; Sober [2008: 32-34]) do, the evidence does not confirm the remaining false hypothesis relative to the true hypothesis.[20]

Whenever the probability that an agent assigns to a false hypothesis goes up, it can potentially lead to bad consequences. For instance, if the increase in probability takes us past the threshold required to convict a suspect, we might execute Dick even though he is innocent. But

---

[18] At least, it represents progress unless 'the Bayesian agent has been so unfortunate as to assign the true hypothesis a zero prior' (Earman [1992: 163]). Since we assume that the hypotheses under consideration are jointly exhaustive, the true hypothesis is included. In addition, in our 'elimination experiment' examples, the agent starts out assigning a non-zero probability to each of the hypotheses.

[19] See Godfrey-Smith [2011: 1294-95] and Fallis [2015: 384-86] for further discussion of why these philosophers have themselves been misled.

[20] If an agent shifts from **s** to **r** on the basis of evidence *e*, then $r_k/s_k = \text{pr}(h_k|e)/\text{pr}(\text{h}_\text{k}) = \text{pr}(e|h_k)/\text{pr}(e)$. So, $r_i/s_i \geq r_j/s_j$ if and only if $\text{pr}(e|h_i) \geq \text{pr}(e|h_j)$. Thus, according to Ian Hacking's 'Law of Likelihoods,' *e* supports h_i at least as much as it does h_j if and only if $r_i/s_i \geq r_j/s_j$.

consideration of such practical matters would take us beyond our focus as epistemologists on the purely *epistemic* betterness of cognitive states.

Any time we get evidence that definitively eliminates a false hypothesis (and provides no information regarding the remaining hypotheses), it is *epistemically* beneficial. The claim that **r** = (1/2, 1/2, 0) is epistemically better than **s** = (1/3, 1/3, 1/3) when h_1 is true is a special case of the following monotonicity principle:

M3. All other things being equal, if **r** assigns a lower probability to some false hypothesis than **s** does, then **r** is epistemically better than **s**.[21]

In fact, it might even be suggested that, any time we get evidence that raises the probability of the true hypothesis in at least as a high a ratio as it raises the probability assigned to any false hypothesis, there is no epistemic downside. That is, we might say that, as long as the evidence confirms the true hypothesis at least as much as it does any false hypothesis, things have not gotten epistemically worse. The claim that **r** = (1/2, 1/2, 0) is epistemically at least as good as **s** = (1/3, 1/3, 1/3) when h_1 is true is a special case of the even stronger monotonicity principle (essentially in terms of likelihoods):

M4. If $r_i/s_i \geq r_j/s_j$ for all j, then $u_i(\mathbf{r}) \geq u_i(\mathbf{s})$.

M4 implies **M1**, **M2**, and **M3**. Also, the logarithmic rule and the spherical rule both endorse this principle.[22]

The Brier rule, however, does not endorse **M3** or **M4**. There are cases where a false hypothesis is definitively eliminated (and the probabilities assigned to the other hypotheses stay in the same ratios), but the Brier rule says that things *have* gotten epistemically worse. For

---

[21] More formally, what **M3** says is that, if $r_j < s_j$ for some $j \neq i$, and there is a real number α such that $\alpha r_k = s_k$ for all $k \neq j$, then $u_i(\mathbf{r}) > u_i(\mathbf{s})$.

[22] This claim is again trivial for the logarithmic rule. A proof of this claim for the spherical rule is in the appendix.

instance, suppose that h_1 is true, **r** = (1/3, 2/3, 0), and **s** = (1/4, 1/2, 1/4). According to the Brier rule, $u_i(\mathbf{r}) = 0.111$ and $u_i(\mathbf{s}) = 0.125$.[23] Moreover, this is not just an isolated case. For all probabilities *a* and *b* such that $b/a > 1.88$, the Brier rule incorrectly says that $\mathbf{s} = (a, b, a)$ is epistemically better than $\mathbf{r} = (a/(a+b), b/(a+b), 0)$ when h_1 is true. Thus, the Brier rule provides an incorrect epistemic ordering of basic cognitive states, and it should be rejected as a measure of epistemic utility.

8. Objections and Replies

8.1 *The Lockean Thesis*

According to the 'Lockean Thesis,' you *categorically believe* a hypothesis if you assign a 'sufficiently high' probability to that hypothesis (see Foley [1993: 140]). With this in mind, it might be suggested that **s** is epistemically better than **r** when h_1 is true (as the Brier rule says) because the shift from 1/2 to 2/3 in the probability assigned to h_2 takes an agent from suspending judgment on a falsehood to outright believing a falsehood. However, it is not clear exactly where the threshold for categorical belief lies (e.g., that it lies between 1/2 and 2/3). As Richard Foley [1993: 142] notes, 'there doesn't seem to be a non-arbitrary way of identifying even a vague threshold.' Also, it is not clear that categorical belief in a falsehood represents an epistemic cost over-and-above whatever happens to the agent's credences.

[23] As noted above, all other things being equal, epistemic utility on the Brier rule decreases as the total probability assigned to the false hypotheses is concentrated on fewer false hypotheses. In many cases, including our 'elimination experiment' counter-example, this 'epistemic cost' is (according to the Brier rule) enough to outweigh the epistemic benefit of an increase in the probability assigned to the true hypothesis. Knab and Schoenfield [2015] give a different example that also displays this effect. In offering an analysis of *misleading evidence*, Lewis and Fallis [2014] independently discuss a very similar case. However, these examples are not as obviously in conflict with scientific practice as our 'elimination experiment' counter-example.

In any event, there are 'elimination experiment' counter-examples to the Brier rule that definitely do not involve exceeding the threshold for categorical belief in a falsehood. According to Foley [1993, 142], 'we will want to stipulate that for belief you need to have more confidence in a proposition than its negation.' But the Brier rule says that **s** = (3/20, 8/20, 3/20, 3/20, 3/20) is epistemically better than **r** = (3/17, 8/17, 3/17, 3/17, 0) when h_1 is true even though the probability assigned to each false hypothesis is less than 1/2 in both cases.

8.2 *Expected epistemic utility*

For any proper scoring rule, including the Brier rule, the *expected* epistemic utility of **r** = (1/3, 2/3, 0) *is* greater than the *expected* epistemic utility of **s** = (1/4, 1/2, 1/4) after one gets evidence that definitively eliminates h_3. But the problem here is that the Brier rule says that the *actual* epistemic utility of **r** is lower than the *actual* epistemic utility of **s** given that h_1 is true.

Admittedly, the results of experiments are sometimes misleading such that actual epistemic utility goes down even though expected epistemic utility goes up. But there is nothing at all misleading about a result that definitively eliminates a false hypothesis (and that has no other effect on one's cognitive state). As noted above, ruling out false possibilities is clear scientific progress. For instance, when Sam eliminates Harry as a suspect, her cognitive state improves (in epistemic terms), regardless of whether Tom or Dick is the guilty party.

In order to press this objection, defenders of the Brier rule would need to explain why simply eliminating a false hypothesis would decrease *actual* epistemic utility. In other words, they need to have a story about what is epistemically bad about 'elimination experiments.' It is not enough to merely rely on the fact that 'elimination experiments' increase *expected* epistemic utility according to the Brier rule. The expected epistemic utility of performing an experiment is a weighted average of the actual epistemic utility at the various possible worlds (i.e., the world where h_1 is true, the world where h_2 is true, etc.). Thus, within the project of epistemic consequentialism, actual epistemic utility is conceptually prior to expected epistemic utility. The fact that a function preserves our intuitions about epistemic utility at the higher level of expected epistemic utility is irrelevant if it violates our intuitions about epistemic utility at a more basic level.

8.3 *Context sensitivity*

Even if one accepts that getting evidence that definitively eliminates a false hypothesis is always epistemically beneficial, one might deny that there is a fact of the matter about whether **r** = (1/3, 2/3, 0) is epistemically better than **s** = (1/4, 1/2, 1/4) when h_1 is true. Although it is epistemically beneficial for an agent to shift from **s** to **r** in the context of an 'elimination experiment,' there might be other contexts in which it is not epistemically beneficial.

However, claiming that there is an incommensurability in epistemic betterness here would be to give up on the project of epistemic consequentialism that these epistemologists are working within. We would have hoped that at least part of the reason why it is epistemically good to definitively eliminate a false hypothesis (even if you learn nothing about the other hypotheses) is that it puts the agent into an epistemically better cognitive state. In any event, since everyone must agree that there are some contexts in which shifting from **s** = (1/4, 1/2, 1/4) to **r** = (1/3, 2/3, 0) when h_1 is true is epistemically beneficial, we would not want to insist (as the Brier rule does) that **r** is *always* epistemically worse than **s**.

8.4 *Diachronic updating*

We have motivated the claim that $\mathbf{r}$ = (1/3, 2/3, 0) is epistemically better than $\mathbf{s}$ = (1/4, 1/2, 1/4) when h_1 is true by appealing to a story about diachronic updating. Thus, one might worry that we have illicitly shifted the focus from synchronic issues to diachronic issues in midstream. However, our appeal to diachronic updating is only used to show that the Brier rule orders basic cognitive states incorrectly. Formal epistemologists adopt essentially the same strategy in order to show that measures of epistemic utility must be proper scoring rules. In any event, the claim that $\mathbf{r}$ is epistemically better than $\mathbf{s}$ when h_1 is true can also be motivated in other ways. For instance, consider two agents whose cognitive states only differ in that one agent knows that h_3 is definitely not true and the other agent does not. It seems clear that the first agent's cognitive state is epistemically better than the second agent's.

8.5 *Measures of confirmation*

As noted above, the law of likelihoods is an especially attractive way to compare the amount of evidential support that a piece of evidence provides to competing hypotheses. But it is certainly not the only possible way to do so. For instance, instead of taking the degree to which e confirms h_i to be a function of $\mathrm{pr}(h_i|e)$ / $\mathrm{pr}(h_i)$, we might take it to be a function of $\mathrm{pr}(h_i|e) - \mathrm{pr}(h_i)$ or a function of $\mathrm{pr}(e|h_i)$ / $\mathrm{pr}(e|{\sim}h_i)$ (see Sober [2008: 16]). According to the 'difference' measure and the 'likelihood ratio' measure, a piece of evidence that takes an agent's cognitive state from $\mathbf{s}$ = (1/4, 1/2, 1/4) to $\mathbf{r}$ = (1/3, 2/3, 0) confirms h_2 to a greater degree than it confirms h_1. This would seem to provide support for the verdict of the Brier rule that $\mathbf{s}$ is epistemically better than $\mathbf{r}$ if h_1 is true. Unfortunately though, these other measures do not vindicate *all* of the verdicts of the Brier rule. For instance, the Brier rule says that $\mathbf{r}$ = (3/7, 4/7, 0) is epistemically better than $\mathbf{s}$ = (3/10, 4/10, 3/10) if h_1 is true even though $r_2 - s_2 > r_1 - s_1$. (The likelihood ratio measure does not vindicate the verdict of the Brier rule in this case either.) In any event, it would be strange to let the choice of a measure of epistemic utility prejudge the choice of a measure of confirmation. And it would be especially strange for defenders of the Brier rule (such as Joyce [2004]) who are pluralists about measures of confirmation.

8.6 *Boolean algebras*

Instead of simply taking credences over a partition of possible hypotheses as we have done here, some formal epistemologists (e.g., Joyce [2009: 263]) take credences over a Boolean algebra of possible hypotheses. In other words, they explicitly consider the probability that Tom or Dick is guilty as well as the probability that Tom is guilty and the probability that Dick is guilty.[24] This choice changes the magnitude of the epistemic utility that the Brier rule assigns to cognitive states, but it does not change the ordering of those states (at least if we continue to assume that an agent has coherent credences). So, the Brier rule is still subject to our 'elimination experiment' counter-example.

Admittedly, if we consider the shift from **s** = (1/4, 1/2, 1/4) to **r** = (1/3, 2/3, 0) when h_1 is true in the context of a full Boolean algebra, the probability assigned to the false hypothesis h_2 goes up in a higher ratio than the probability assigned to the true hypothesis $h_1 \vee h_3$. Indeed, the probability assigned to the true hypothesis $h_1 \vee h_3$ actually goes down (from 1/2 to 1/3). This shows that the monotonicity principles (such as **M4**) do not translate directly to the context of a full Boolean algebra.[25] But it does not show that our 'elimination experiment' counter-example

[24] A full Boolean Algebra includes conjunctions, as well as disjunctions, of the basic hypotheses. However, since we are assuming here that the hypotheses under consideration are mutually exclusive, any conjunctions must be assigned a probability of zero.

[25] It is possible to modify some of these principles so that they are applicable to a full Boolean algebra of hypotheses in which more than one hypothesis can be true. For instance, we could

to the Brier rule is not still a counter-example when we use a full Boolean algebra rather than a partition.

When we use a full Boolean algebra, the probability assigned to some false hypothesis will often go up in a higher ratio than the probability assigned to some true hypothesis even when things have clearly gotten epistemically better. For instance, if an agent's cognitive state goes from **s** = (1/3, 1/3, 1/3) to **r** = (1/2, 1/2, 0) when h_1 is true, the probability assigned to h_2 goes up and the probability assigned to $h_1 \vee h_3$ goes down (from 2/3 to 1/2). Nevertheless, all proper scoring rules count this shift as an epistemic improvement (and this verdict does not depend on our choice of confirmation measure).

What matters for epistemic improvement is how the probabilities assigned to the basic hypotheses that make up the partition (not the probabilities assigned to all of the hypotheses in the Boolean algebra) change relative to each other. When Sam eliminates Harry as a suspect, it is not that *on balance* her epistemic utility goes up, despite the fact that her credence in the true proposition 'Tom or Harry did it' goes down. Rather, we should ignore Sam's credences in the disjunctions, and simply evaluate her epistemic utility based on her credence that Tom did it, her credence that Dick did it, and her credence that Harry did it.

It should be noted that the defenders of the Brier rule cannot accuse us here of adopting an inappropriately coarse-grained representation of an agent's cognitive state. We are not starting with a full Boolean algebra and then arbitrarily choosing from it a partition of mutually exclusive and jointly exhaustive hypotheses. Instead, we are starting with a partition of mutually

---

rewrite **M4** as **M4***: If $r_i/s_i \geq r_j/s_j$ for all i, j such that h_i is true and h_j is false, then **r** is epistemically at least as good as **s**. Although this principle true, it is far too weak for our purposes. Elimination experiments are always epistemically beneficial, but **M4*** does not entail this.

exclusive and jointly exhaustive hypotheses from which a full Boolean algebra can be generated. Thus, the partition is as fine-grained as the corresponding Boolean algebra.

8.7 *Conditionalization*

It might be suggested that the real problem is with simple conditionalization rather than with the Brier rule. That is, when we learn for sure that h_3 is false, we should not conditionalize on this evidence. So, the fact that definitively eliminating a false hypothesis is always epistemically beneficial does not tell us anything about whether **r** = (1/3, 2/3, 0) is epistemically better than **s** = (1/4, 1/2, 1/4). But this seems like a rather radical way for epistemologists to try to rescue the Brier rule as a measure of epistemic utility. It would seem particularly odd given that, as noted above, the Brier rule is typically used to *defend* the epistemic value of conditionalization.

8.8 *Positives versus negatives*

Finally, it might be suggested that, even if the 'elimination experiment' counter-examples count against the Brier rule, this negative must be weighed against the positives of the Brier rule.[26] Most notably, the fact that it vindicates probabilism and conditionalization counts strongly in favor of the Brier rule as a measure of epistemic utility.

However, formal epistemologists presumably do not want a measure of epistemic utility that flies in the face of widely accepted norms of scientific practice (see Maher [1993: 209-16]; Fallis [2007: 218]). Jumping to a conclusion in the absence of any evidence is clearly a bad thing epistemically. Thus, if a proposed measure of epistemic utility lacks propriety, that is a deal breaker. In a similar vein, all other things being equal, definitively eliminating a false

[26] Thanks to an anonymous referee for suggesting this sort of objection.

hypothesis is clearly a good thing epistemically. Thus, it is disqualifying if a proposed measure of epistemic utility says otherwise.

In any event, vindicating probabilism and conditionalization does not count *uniquely* in favor of the Brier rule. Although he utilized the Brier rule in his earlier work, Richard Pettigrew [2013: 904-05] has subsequently shown how to derive these two results using weaker assumptions about how to measure epistemic utility.[27] Thus, many of the exciting new results in the project of epistemic consequentialism do not rely on taking the Brier rule as our measure of epistemic utility.

9. Conclusion

The Brier rule is the most popular choice (by far) among formal epistemologists for a measure of epistemic utility. However, the Brier rule is clearly *not* a good measure of epistemic utility. It is often wrong about whether one cognitive state is epistemically better than another. In particular, it incorrectly says that **s** = (1/4, 1/2, 1/4) is epistemically better than **r** = (1/3, 2/3, 0) when h_1 is true. Contrary to what the Brier rule says, it seems clear that definitively eliminating a false hypothesis is always epistemically beneficial. So, given that there are many other candidates

[27] The standard versions of the logarithmic rule and the spherical rule (as stated above) do not vindicate probabilism. Joyce [2009, 275] proposes alternative versions of these rules that do. Unfortunately, unlike the standard versions, Joyce's versions of these rules are (just like the Brier rule) subject to 'elimination experiment' counter-examples. (His version of the logarithmic rule does say that **r** = (1/3, 2/3, 0) is epistemically better than **s** = (1/4, 1/2, 1/4) when h_1 is true, but it says that **s** = (1/10, 8/10, 1/10) is epistemically better than **r** = (1/9, 8/9, 0) when h_1 is true.) So, vindicating probabilism while avoiding 'elimination experiment' counter-examples might not be trivial.

available, epistemologists can and should stop using and defending the Brier rule as a measure of epistemic utility.[28]

University of Arizona

University of Miami

References

Bickel, J. E. 2007. Some Comparisons Among Quadratic, Spherical, and Logarithmic Scoring Rules, *Decision Analysis* 4/2: 49-65.

Bruner, Justin P. 2015. Disclosure and Information Transfer in Signaling Games, *Philosophy of Science* 82/4: 649-66.

Earman, John 1992. *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*, Cambridge: MIT Press.

Fallis, Don 2007. Attitudes Toward Epistemic Risk and the Value of Experiments, *Studia Logica* 86/2: 215-46.

Fallis, Don 2015. Skyrms on the Possibility of Universal Deception, *Philosophical Studies* 172/2: 375-97.

Fallis, Don and Gerrard Liddell 2002. Further Results on Inquiry and Truth Possession, *Statistics and Probability Letters* 60/2: 169-82.

Fallis, Don and Dennis Whitcomb 2009. Epistemic Values and Information Management, *The Information Society* 25/3: 175-89.

Foley, Richard 1993. *Working Without a Net: A Study of Egocentric Epistemology*, New York: Oxford University Press.

Godfrey-Smith, Peter 2011. Review of *Signals* by Brian Skyrms, *Mind* 120/480: 1288-97.

[28] We would like to thank Kobus Barnard, David Black, Kenny Easwaran, Branden Fitelson, Will Fleisher, Martin Frické, Peter Godfrey-Smith, Simon Goldstein, Stephen Hetherington, Terry Horgan, Jenann Ismael, James Joyce, Gerrard Liddell, Kay Mathiesen, Richard Pettigrew, Daniel Rubio, Jonah Schupbach, Elliott Sober, Julia Staffel, Dan Zelinski, and two anonymous referees for helpful feedback on earlier versions of this material. Much of this work was carried out while the first author was a visiting fellow at the Tanner Humanities Center at the University of Utah.

Goldman, Alvin I. 1999. *Knowledge in a Social World*, Oxford: Oxford University Press.

Greaves, Hilary and David Wallace 2006. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility, *Mind* 115/459: 607-32.

Hacking, Ian 1965. *Logic of Statistical Inference*, Cambridge: Cambridge University Press.

Joyce, James M. 2004. On the Plurality of Probabilist Measures of Evidential Relevance, Paper presented at the meeting of the Philosophy of Science Association.

Joyce, James M. 2009. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief, in *Degrees of Belief*, ed. F. Huber and C. Schmidt-Petri, Dordrecht: Springer: 263-97.

Kelly, Thomas 2003. Epistemic Rationality as Instrumental Rationality: A Critique, *Philosophy and Phenomenological Research* 66/3: 612-40.

Kierland, Brian, and Bradley Monton 2005. Minimizing Inaccuracy for Self-locating Beliefs, *Philosophy and Phenomenological Research* 70/2: 384-95.

Knab, Brian and Miriam Schoenfield 2015. A Strange Thing about the Brier Score, *M-Phi*, http://m-phi.blogspot.nl/2015/03/a-strange-thing-about-brier-score.html

Leitgeb, Hannes, and Richard Pettigrew 2010a. An Objective Justification of Bayesianism I: Measuring Inaccuracy, *Philosophy of Science* 77/2: 201-35.

Leitgeb, Hannes, and Richard Pettigrew 2010b. An Objective Justification of Bayesianism II: The Consequences of Measuring Inaccuracy, *Philosophy of Science* 77/2: 236-72.

Levinstein, Benjamin Anders 2012. Leitgeb and Pettigrew on Accuracy and Updating, *Philosophy of Science* 79/3: 413-24.

Lewis, Peter and Don Fallis 2014. Misleading Evidence, URL = <http://ssrn.com/abstract=2462771>

Maher, Patrick 1990. Why Scientists Gather Evidence, *British Journal for the Philosophy of Science* 41/1: 103-19.

Maher, Patrick 1993. *Betting on Theories*, Cambridge: Cambridge University Press.

Martínez, Manolo 2015. Deception in Sender-Receiver Games, *Erkenntnis* 80/1: 215-27.

Oddie, Graham 1997. Conditionalization, Cogency, and Cognitive Value, *British Journal for the Philosophy of Science* 48/4: 533-41.

Oddie, Graham 2014. Truthlikeness, *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/entries/truthlikeness/>

Pettigrew, Richard 2013. Epistemic Utility and Norms for Credence, *Philosophy Compass* 8/10: 897-908.

Skyrms, Brian 2010. *Signals: Evolution, Learning, & Information*, Oxford: Oxford University Press.

Smead, Rory 2014. Deception and the Evolution of Plasticity, *Philosophy of Science* 81/5: 852-65.

Sober, Elliott 2008. *Evidence and Evolution: The Logic Behind the Science*, Cambridge: Cambridge University Press.

Appendix

Proof of M2 for the Brier rule

Suppose $r_i > s_i$, and $r_j \leq s_j$ for all $j \neq i$. According to the Brier rule,

$$u_i(\mathbf{s}) = 2s_i - \sum_k s_k^2. \tag{1}$$

Extracting the term s_i^2 from the sum on the right yields

$$u_i(\mathbf{s}) = (2s_i - s_i^2) - \sum_{j\neq i} s_j^2, \tag{2}$$

and by the same reasoning

$$u_i(\mathbf{r}) = (2r_i - r_i^2) - \sum_{j\neq i} r_j^2 \tag{3}$$

But since $r_i > s_i$, $0 \leq r_i \leq 1$ and $0 \leq s_i \leq 1$,

$$(2r_i - r_i^2) > (2s_i - s_i^2). \tag{4}$$

Furthermore, since $r_j \leq s_j$ for all $j \neq i$,

$$\sum_{j\neq i} r_j^{\mathbf{2}} \leq \sum_{j\neq i} s_j^2. \tag{5}$$

That is, the first term on the right in (3) is larger than the first term in (2), and the second term in (3) is no larger than the second term in (2). Hence $u_i(\mathbf{r}) > u_i(\mathbf{s})$.

Proof of M2 for the spherical rule

Suppose $r_i > s_i$, and $r_j \leq s_j$ for all $j \neq i$. According to the spherical rule,

$$u_i(\mathbf{s}) = \frac{s_i}{\sqrt{\sum_k s_k^2}}. \tag{6}$$

Inverting and squaring gives

$$\frac{1}{(u_i(\mathbf{s}))^2} = \frac{\sum_k s_k^2}{s_i^2}, \tag{7}$$

and extracting the term s_i^2 from the sum on the right yields

$$\frac{1}{(u_i(\mathbf{s}))^2} = \frac{s_i^2 + \sum_{j \neq i} s_j^2}{s_i^2} = 1 + \frac{\sum_{j \neq i} s_j^2}{s_i^2}. \tag{8}$$

Similarly,

$$\frac{1}{(u_i(\mathbf{r}))^2} = 1 + \frac{\sum_{j \neq i} r_j^2}{r_i^2}. \tag{9}$$

Since $r_i > s_i$, the denominator of the fraction on the right in (9) is greater than the corresponding denominator in (8), and since $r_j \leq s_j$ for all $j \neq i$, the numerator in (9) is no greater than the numerator in (8). Hence

$$\frac{1}{(u_i(\mathbf{s}))^2} > \frac{1}{(u_i(\mathbf{r}))^2} \tag{10}$$

and so $u_i(\mathbf{r}) > u_i(\mathbf{s})$.

Proof of M4 for the spherical rule

Suppose $r_i/s_i \geq r_j/s_j$ for all j. That is, suppose each r_j can be written as $x_j s_j$, where $x_i \geq x_j$ for all j, and the $x_j s_j$ sum to 1. According to the spherical rule,

$$u_i(\mathbf{s}) = \frac{s_i}{\sqrt{\sum_j s_j^2}}. \tag{11}$$

Similarly, since $r_j = x_j s_j$ for all j,

$$u_i(\mathbf{r}) = \frac{r_i}{\sqrt{\sum_j r_j^2}} = \frac{x_i s_i}{\sqrt{\sum_j x_j^2 s_j^2}}. \tag{12}$$

Dividing top and bottom by x_i gives

$$u_i(\mathbf{r}) = \frac{s_i}{\sqrt{\sum_j \frac{x_j^2}{x_i^2} s_j^2}}. \tag{13}$$

Since $x_i \geq x_j$ for all j, the denominator in (13) is smaller than the denominator in (11), or equal in the special case in which the x_j are all equal (to 1). Hence $u_i(\mathbf{r}) \geq u_i(\mathbf{s})$.