Noname manuscript No.
(will be inserted by the editor)

Accuracy, conditionalization, and probabilism

Peter J. Lewis & Don Fallis

the date of receipt and acceptance should be inserted later

Abstract Accuracy-based arguments for conditionalization and probabilism appear to have a significant advantage over their Dutch Book rivals. They rely only on the plausible epistemic norm that one should try to decrease the inaccuracy of one's beliefs. Furthermore, conditionalization and probabilism apparently follow from a wide range of measures of inaccuracy. However, we argue that there is an under-appreciated diachronic constraint on measures of inaccuracy which limits the measures from which one can prove conditionalization, and none of the remaining measures allow one to prove probabilism. That is, among the measures in the literature, there are some from which one can prove conditionalization, others from which one can prove probabilism, but none from which one can prove both. Hence at present, the accuracy-based approach cannot underwrite both conditionalization and probabilism.

A central concern of epistemology is uncovering the rational constraints on an agent's credences, both at a time and over time. At a time, it is typically maintained that an agent's credences should conform to the probability axioms, and over time, it is often maintained that an agent's credences should conform to conditionalization. How could such norms be justified? The traditional approach is to show that if your credences violate these norms, then there is a set of bets, each of which you consider fair, but which collectively are such that if you accept them all you will lose money whatever happens. Since you do not want to be a "money pump", you should adopt coherent credences and you should conditionalize. However, this *Dutch book* strategy rests on controversial assumptions concerning prudential rationality and its connection to epistemic rationality.

Address(es) of author(s) should be given

The prudential elements may not be essential to the Dutch book approach (Vineberg 2012). But even so, it would be better to be able to derive probabilism and conditionalization from a basic norm that is clearly epistemic.[1] A more recent approach seeks to do precisely that: to derive probabilism and conditionalization from the intuitive epistemic norm that you should endeavor to make your credences as accurate—as close to the truth—as possible. Drawing on the work of Joyce (1998; 2009), Greaves and Wallace (2006) and Predd et al. (2009), Pettigrew (2013) argues that the accuracy-based approach vindicates both probabilism and conditionalization. We argue that this conclusion is too strong: at present, the accuracy-based approach can at best vindicate *either* conditionalization *or* probabilism, but not both.

Our argument turns on the features of various proposed measures of accuracy. The accuracy-based approach is predicated on the assumption that the accuracy of your credences can be measured. Pettigrew (2013, 905) argues that it is a strength of the accuracy-based approach that conditionalization and probabilism follow from a wide range of measures, so that it doesn't matter which measure is used to assess the accuracy of an agent's credences. Our counter-argument is that of the measures in the literature, some vindicate conditionalization, and some vindicate probabilism, but there is no measure from which both conditionalization and probabilism can be derived.

In section 1 we briefly rehearse the accuracy based arguments for conditionalization and probabilism. In section 2 we locate our critique of these arguments among the existing literature. In section 3 we introduce a diachronic constraint, Elimination, that (we argue) any accuracy measure must obey if it is to ground conditionalization.[2] In section 4 we present our main argument, which goes as follows. Three well-known rules for measuring the accuracy of an agent's credences (Brier, log and spherical) allow the derivation of conditionalization, but one of these (Brier) violates Elimination. The remaining rules must be generalized if they are to be used to derive probabilism, but the obvious generalizations all violate Elimination. Hence we conclude that none of these popular rules can be used to prove both conditionalization and probabilism (and if there is a synchronic analog of Elimination, then none of them can be used to prove probabilism at all). In section 5 we respond to objections to Elimination. In section 6 we assess the prospects for finding some *new* accuracy measure that could ground both conditionalization and probabilism.

1 Accuracy: the state of the art

First, let us briefly run through the arguments via which conditionalization and probabilism are claimed to follow from considerations of accuracy, starting

[1] Also, Dutch book arguments arguably rest on substantial assumptions about your credences in propositions and their negations (Hedden 2013).

[2] We introduce the Elimination constraint in Fallis and Lewis (2016). The current paper goes beyond our prior work in assessing the consequences of Elimination for proofs of probabilism and conditionalization.

with conditionalization. Suppose you have credences $c(X_i) = b_i$; that is, you have credences $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in propositions $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Suppose further that the propositions form a partition, i.e. they are exhaustive and mutually exclusive, so that exactly one of them is true. The accuracy approach takes it that your primary epistemic goal is having credences that are as accurate as possible, where complete accuracy is a credence of 1 in the true proposition and a credence of 0 in each of the false propositions. The closer your credences are to complete accuracy, the better.

For this epistemic goal to make sense, we need a measure of closeness. In what follows we will discuss several such measures, expressed as measures of *inaccuracy*: $I_i(\mathbf{b})$ is the inaccuracy of credences $\mathbf{b}$ when proposition X_i is true. The larger $I_i(\mathbf{b})$, the further your credences are from the truth; hence your goal is to minimize the value of $I_i(\mathbf{b})$. Obviously not just any function of your credences makes for a reasonable inaccuracy measure. One constraint that is usually considered essential is that any such measure must obey Strict Propriety:

Strict Propriety: For any distinct probabilistic credences $\mathbf{b}$ and $\mathbf{b}'$, $\sum_i b_i I_i(\mathbf{b}) < \sum_i b_i I_i(\mathbf{b}')$.

Strict Propriety says that the expected inaccuracy of your current credences $\mathbf{b}$ is lower than the expected inaccuracy of any alternative credences $\mathbf{b}'$ you might adopt, where the expectation is calculated according to your current credences. If it fails, then the injunction to minimize inaccuracy makes your beliefs pathologically unstable: you can lower your expected inaccuracy by shifting your credences, even in the absence of new evidence. Hence Strict Propriety serves as a reasonable constraint on measures of inaccuracy.[3]

Even given this constraint, though, there is still an infinite variety of strictly proper inaccuracy measures: Joyce (2009, 277) provides a general recipe for constructing them. We return to consider the full variety of measures in section 6; for present purposes, it will suffice to consider a few prominent examples. The most frequently cited measures in the literature are the Brier rule (or quadratic rule), the log rule, and the spherical rule:

Simple Brier rule: $I_i(\mathbf{b}) = (1 - b_i)^2 + \sum_{j \neq i} b_j^2$.
Simple log rule: $I_i(\mathbf{b}) = -\ln b_i$.
Simple spherical rule: $I_i(\mathbf{b}) = 1 - b_i / \sqrt{\sum_j b_j^2}$.

They are all strictly proper. We call these rules *simple* to distinguish the versions applicable to a partition, defined here, from the versions applicable to a Boolean algebra, defined later. By far the dominant measure in the literature is the Brier rule: the Brier rule has been defended by epistemologists (Joyce 2009, 290; Leitgeb and Pettigrew 2010, 219; Pettigrew 2016, 67), and is frequently cited as the prime example of an inaccuracy measure (Greaves and Wallace 2006, 627; Pettigrew 2013, 899).

[3] See Maher (1990, 112). However, as discussed in section 2, Blackwell and Drucker (2019) challenge the reasonableness of Strict Propriety.

Suppose you obtain new evidence E; how should you redistribute your credence over the propositions $\mathbf{X}$? According to conditionalization, your new credences c' should be your old credences conditional on E: $c'(X_i) = c(X_i|E)$. Can conditionalization be justified by appeal to accuracy? If your goal is to minimize your inaccuracy, presumably the best you can do is to minimize your *expected* inaccuracy given your prior credences $\mathbf{b}$. Greaves and Wallace (2006) prove that conditionalization minimizes expected inaccuracy for any measure of inaccuracy $I_i(\mathbf{b})$ satisfying Strict Propriety.

Now let us turn to the argument that your credences at a time should obey the probability axioms. So far, we have been assuming that the propositions we are interested in form a partition. But the probability axioms include constraints on your credences in disjunctions, and to model such constraints we need to allow that more than one of the propositions you are considering can be true. To that end, suppose that you have credences $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in propositions $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, where the set of propositions forms a Boolean algebra, i.e. it is closed under negation and disjunction. So now we can no longer model a possible world simply as an index (picking out the unique true proposition); instead, we need to label each proposition separately as either true or false. That is, a possible world is specified by $\omega = (\omega_1, \omega_2, \ldots \omega_n)$, where $\omega_i = 1$ when X_i is true and $\omega_i = 0$ when X_i is false. The simple Brier, simple log, and simple spherical rules can be generalized to a apply to a Boolean algebra as follows:

Symmetric Brier rule: $I(\omega, \mathbf{b}) = \sum_i (b_i - \omega_i)^2$.
Symmetric log rule: $I(\omega, \mathbf{b}) = \sum_i -\ln |(1 - \omega_i) - b_i|$.
Symmetric spherical rule: $I(\omega, \mathbf{b}) = \sum_i 1 - \frac{|(1-\omega_i)-b_i|}{\sqrt{b_i^2+(1-b_i)^2}}$.

We call these versions *symmetric* to distinguish them from the *simple* versions, and to draw attention to a certain property they have: they treat truth and falsity symmetrically, in the sense that inaccuracy is always the same function of the distance between each credence and the truth value of the corresponding proposition, regardless of whether that proposition is true or false.[4] This property will be important later.

The general strategy for defending probabilism based on accuracy goes as follows. Suppose that your current credences are incoherent—that is, they violate the probability axioms. Then one can appeal to a measure of inaccuracy to show that there are coherent credences that *dominate* your current credences—that are more accurate than your current credences whatever the truth values of the propositions concerned. If your goal is to minimize inaccuracy, this gives you a clear reason to avoid incoherent credences: there are always coherent credences that are more accurate, whatever the world is like.

Predd et al. (2009) adopt this proof strategy. Their proof relies on two assumptions. The first is Additivity:

[4] There are many different notions of symmetry in the accuracy literature; this is the one Joyce (2009, 274) calls "0/1-symmetry".

Additivity: $I(\omega, \mathbf{b})$ can be expressed as $\sum_i s(\omega_i, b_i)$, where s is a continuous function of your credence in proposition X_i and its truth value.

Additivity states that the inaccuracy of your beliefs in a set of propositions is just the sum of your inaccuracies in the propositions taken individually—that is, $s(\omega_i, b_i)$ is the inaccuracy of your belief in proposition X_i, and $I(\omega, \mathbf{b})$ is just the sum of these inaccuracies for all the propositions you are considering. Note that it also contains the requirement that the inaccuracy measure should be continuous. The symmetric Brier, log and spherical rules are obviously additive, since each is expressed as a sum over propositions, and each uses a continuous function.

The second assumption is Strict Propriety. For an additive inaccuracy measure, Strict Propriety can be expressed in terms of your inaccuracy function for a single proposition $s(b_i, \omega_i)$ as follows:

Strict Propriety (for an additive measure): $b_i s(x, 1)+(1-b_i)s(x, 0)$ is uniquely minimized at $x = b_i$.

The symmetric Brier, log and spherical rules, like their simple counterparts, are strictly proper. Predd et al. (2009) prove that any additive, strictly proper inaccuracy measure entails probabilism—that is, for any incoherent set of credences $\mathbf{b}$, there is a coherent set $\mathbf{b}^*$ that is less inaccurate than $\mathbf{b}$ in every possible world.

As Pettigrew (2013, 905) notes, Greaves and Wallace (2006) and Predd et al. (2009) prove strong results. Any inaccuracy measure satisfying Strict Propriety can be used to vindicate conditionalization, and Strict Propriety is a constraint we would expect a reasonable inaccuracy measure to obey anyway. Any inaccuracy measure satisfying Strict Propriety and Additivity can be used to vindicate probabilism, and while Additivity is perhaps not forced on us in the way that Strict Propriety appears to be, it is certainly intuitive.[5] As we have seen, there are several available measures satisfying Additivity and Strict Propriety, so it initially looks like the accuracy-based program can justify both probabilism and conditionalization based on minimal premises. Our purpose in this paper is to argue that matters are not so straightforward. But first, let us consider some prior critiques of the program, to establish that they don't make our argument otiose.

2 Critiques and caveats

The accuracy-based program has already come under attack from a number of different directions. Carr (2017) argues that in cases where your credence in a proposition affects its chance—for example, if your confidence that you can perform a handstand affects the chance that you can perform a handstand—conditionalization does not maximize expected accuracy. Talbot (2017) argues that the accuracy-based program entails the repugnant conclusion that

[5] However, Additivity is not beyond question, as Marxen and Rothfus (2018, 317) point out.

minimally accurate credences in a large number of propositions are better than highly accurate credences in a smaller number of propositions. Solving this problem, he argues, requires a distinction between interesting and boring propositions, but such a distinction threatens the accuracy-based argument for probabilism. Blackwell and Drucker (2019) argue that Strict Propriety cannot be motivated in general, and admits counterexamples: for example, if an agent realizes that she has made a mistake in updating her credences in the past, then Strict Propriety forbids her from correcting that mistake by unilaterally shifting her credences in the absence of new evidence.[6]

We do not wish to dismiss these critiques; they are important challenges to accuracy-based epistemology, at least if the goal of epistemology is to provide a global account of the totality of an agent's beliefs. However, their force can be mitigated by focusing on a more modest epistemic project. Consider a particular epistemic inquiry, defined by a particular set of propositions in which an agent is interested. By hypothesis, then, all of the propositions are interesting, and Talbot's worry does not arise. As long as the propositions do not include any where credence affects chance, then Carr's worry does not arise. As long as the agent has not made a mistake concerning these propositions, then Blackwell and Drucker's worry does not arise. Most canonical epistemic inquiries—a detective investigating who committed a crime, or a scientist investigating the behavior of part of the natural world—would seem to typically fulfill these desiderata. And it is certainly a significant result if accuracy-based considerations can vindicate probabilism and conditionalization in canonical cases of this kind. These are the kinds of cases we will focus on.

Our contention, then, is that there is a flaw in the accuracy-based program even if we restrict attention to the most favorable kind of case. In that sense, our critique is more thoroughgoing than those of the authors just mentioned. But in another sense, our aim is more conservative. Unlike these authors, we do not argue that maximizing accuracy sometimes requires one to violate a standard epistemic norm; we take probabilism and conditionalization for granted, at least for canonical cases. Rather, we seek to show that some of the measures of inaccuracy used to prove probabilism and conditionalization entail epistemically problematic conclusions, and hence that the derivation of probabilism and conditionalization within the accuracy-based program is problematic.[7] Our argument is based on an underappreciated constraint on inaccuracy measures, introduced in the following section.

3 Elimination cases

The argument for conditionalization restricts inaccuracy measures to those that are strictly proper. Note that Strict Propriety is only a condition on

[6] We consider Oddie's (2017) critique of the accuracy-based program in section 5.

[7] Blackwell and Drucker (2019) engage in the both kinds of critique: they argue that maximizing accuracy sometimes requires one to violate conditionalization, *and* that the accuracy-based derivation of conditionalization is flawed.

expected inaccuracy. But expected inaccuracy is calculated on the basis of the *actual* inaccuracy that the measure in question ascribes to your credences in the various possible worlds. Presumably there are a number of constraints any such measure must obey if it is to measure actual epistemic inaccuracy rather than something else. For example, if one of your credences shifts towards the truth, while your other credences stay the same, then clearly your actual inaccuracy should decrease. We wish to focus on one such constraint.

The constraint can be motivated by thinking about *elimination cases*. Suppose you are considering a set of mutually exclusive and exhaustive propositions, and suppose that your credences are coherent and that you conditionalize on evidence. You acquire some evidence that eliminates one false proposition—your credence in it becomes zero—but is uninformative regarding the other hypotheses—your credences in them remain in the same proportions. How does this affect the accuracy of your credences?

It seems obvious that your beliefs have become more accurate. If you believe that Tom, Dick or Harry might be the murderer (when in fact Tom did it), and you eliminate Harry while learning nothing about Tom or Dick, then you have made epistemic progress towards the truth, or at least away from falsity. It is true that your credence in the false proposition "Dick did it" goes up, but only by the same proportion that your credence in the true proposition "Tom did it" goes up. Any purported measure of inaccuracy that denies this epistemic progress must be mistaken.

We can codify the lesson of this example as a constraint on inaccuracy measures:

Elimination: For coherent credences over a partition, if $\mathbf{b}$ assigns a zero credence to some false proposition to which $\mathbf{b}'$ assigns a non-zero credence, and credences in the remaining propositions stay in the same ratios, then $\mathbf{b}$ is more accurate than $\mathbf{b}'$.

Unfortunately, the simple Brier rule does not obey this constraint. Let X_1 be "Tom did it", X_2 be "Dick did it", and X_3 be "Harry did it", where unknown to you X_1 is true. Suppose that your initial credences in (X_1, X_2, X_3) are $\mathbf{b} = (1/7, 3/7, 3/7)$. Then according to the simple Brier rule, your initial inaccuracy is $54/49 = 1.10$. Now suppose you acquire some evidence that eliminates X_3, but is uninformative regarding X_1 and X_2. That is, your credence in X_3 becomes 0 and your credences in X_1 and X_2 stay in the same proportions, so that your final credences are $\mathbf{b}^* = (1/4, 3/4, 0)$. Then according to the simple Brier rule, your final inaccuracy is $18/16 = 1.13$. That is, the Brier rule erroneously says that the inaccuracy of your beliefs has gone up.

To press the point, suppose you initially consider a number of exhaustive and mutually exclusive hypotheses, where your credence in the true hypothesis is initially 1/3 of your credence in each of the false hypotheses. That is, your credence in the true hypothesis is initially $1/(3N + 1)$ and your credence in each of the false hypotheses is $3/(3N + 1)$, where N is the number of false hypotheses. You acquire evidence that eliminates the false hypotheses one by one until only the true hypothesis remains. Application of the simple Brier rule shows that your inaccuracy at each step of the elimination process is $9n(n+1)/(3n+1)^2$, where n is the number of false hypotheses that have not yet been eliminated. As n decreases, your inaccuracy gradually increases to a maximum at $n = 1$ (when two hypotheses remain), and then decreases suddenly to zero at $n = 0$ (when only one hypothesis remains). Clearly you are steadily converging on the truth, but according to the Brier rule, your credences are becoming more inaccurate, except at the very last step. It is true that at each step your *expected* inaccuracy goes down, but this is beside the point. The question is whether the Brier rule is an adequate measure of your *actual* inaccuracy, and this case suggests that it is not.

Indeed the strongest evidence for Elimination comes from scientific methodology. Elimination of false hypotheses is a large part of scientific epistemology (Earman 1992, 163); some would say that it is the whole of scientific epistemology (Popper 1963, 51; Mayo 1996, 7). Presumably it is rare in any scientific inquiry to eliminate *every* false alternative to a true hypothesis. So provided that the true hypothesis starts off sufficiently implausible compared to the false hypotheses, the above result shows that according to the simple Brier rule, epistemic progress via the elimination of false hypotheses is impossible: eliminating false hypotheses always makes your credences more inaccurate. This strikes us as highly counterintuitive, although we recognize that intuitions about such cases might vary.

Perhaps, though, the appropriate moral here is that your initial credence in each hypothesis should be the same; given this flat prior, the Brier rule entails that eliminating false hypotheses always decreases your inaccuracy.[8] But a flat prior doesn't solve the problem: suppose you get unlucky and encounter some misleading evidence that decreases your credence in the true hypothesis, but leaves your credence in the false hypotheses unchanged. Then according to the simple Brier rule, this initial shift away from the truth precludes any future epistemic progress. This seems unduly pessimistic. Rather, we take the moral to be that Elimination is required by scientific epistemology, and hence that measures like the simple Brier rule that violate it fail as measures of epistemic accuracy, at least in diachronic contexts.

Another possible response to the argument from scientific methodology is that it is consistent with good methodology to be *misled* occasionally. Cases in which conditionalization lowers your expected inaccuracy but raises your actual inaccuracy are certainly possible, even using a perfectly good measure of inaccuracy. Suppose you are not sure whether a coin is fair or biased. If you toss it five times, get five heads, and conditionalize on the results, your expected inaccuracy for the partition {Fair coin, Heads bias, Tails bias} goes down (according to any reasonable measure). Nevertheless, if the coin is actually fair, your actual inaccuracy for this partition goes up. Perhaps the Tom-Dick-Harry case and its $(N+1)$-hypothesis generalization are simply cases like this, in which a researcher following a perfectly good procedure ends up being misled.

8 Pettigrew (2016) argues for a flat prior on accuracy-based grounds.

Our response to this suggestion is that standard cases of misleading evidence are unlike the Tom-Dick-Harry case in two important respects. First, in the coin-toss case, a string of five heads gives you evidence concerning the partition {Fair coin, Heads bias, Tails bias}, but not infallible evidence: the probability of getting five heads in a row is highest if the coin is biased towards heads, but it is non-zero whichever hypothesis is true. This is what makes it possible for you to be misled: you can get five heads in a row, strongly disconfirming a tails bias, even if the coin is biased towards tails. In the Tom-Dick-Harry case, we can assume that the evidence in question *is* infallible, in the sense that the evidence has probability 0 if Harry did it. Then it is not possible to get misleading evidence in the same way as you can for coin-tossing: there is no way to get evidence that disconfirms "Harry did it" when "Harry did it" is true.

Second, we assume that the evidence in question is completely uninformative regarding "Tom did it" and "Dick did it"; perhaps the evidence has probability 1 whichever of these hypotheses is true. Since the evidence does not differentially confirm one over the other, it cannot thereby be misleading. Again, this is unlike the coin-toss case, in which five heads in a row also provides evidence distinguishing the fair-coin and heads-bias hypotheses, and hence can be misleading, e.g., if fair-coin is disconfirmed relative to heads-bias when the coin is fair. We fail to see how the evidence as described can be misleading in the Tom-Dick-Harry case; at least, the burden seems to be on our opponents to explain the sense in which you are misled. We conclude that elimination cases always decrease inaccuracy, and that any measure that does not concur fails to measure the actual inaccuracy of your beliefs.

4 Measures of inaccuracy

We have argued that for a measure to genuinely measure the actual inaccuracy of your beliefs, at least in contexts where comparison between inaccuracy at distinct times is relevant, it should not be susceptible to elimination counterexamples. Clearly the derivation of conditionalization is one such context. So any measure of inaccuracy that can serve as the basis of a proof of conditionalization should obey Elimination. The simple Brier rule violates Elimination, and hence cannot serve as the basis for a proof of conditionalization.[9]

Fortunately, though, there are alternative inaccuracy measures for partitions we can appeal to. Both the simple log rule and the simple spherical rule satisfy Elimination, and hence neither is susceptible to elimination counterexamples.[10] Hence each can plausibly be claimed to measure epistemic inaccuracy in contexts where comparisons between accuracy at distinct times is

[9] One might reasonably think that acceptable measures of accuracy should obey a stronger principle than Elimination; see Fallis and Lewis (2016).

[10] This is trivial for the log rule, and easily proven for the spherical rule. See Fallis and Lewis (2016).

relevant. Furthermore, each is strictly proper, and so each can be used to underwrite conditionalization via the Greaves and Wallace argument strategy. So there are some inaccuracy measures that vindicate conditionalization, but not all strictly proper measures do so. In particular, the simple Brier rule cannot be used to vindicate conditionalization.

But what about probabilism? There are two considerations to bear in mind here: first that the proof of probabilism requires inaccuracy measures defined over a Boolean algebra, and second that probabilism is a synchronic rather than a diachronic constraint on credences. Regarding the first consideration, when we generalize the inaccuracy measures to a Boolean algebra, we find that all three rules—the symmetric Brier, log and spherical rules—are subject to elimination counterexamples. For the symmetric Brier rule, the counterexample is the same as before, since the symmetric Brier rule directly reduces to the simple Brier rule when applied to a partition.[11] That is, consider a credence shift from $\mathbf{b} = (1/7, 3/7, 3/7)$ to $\mathbf{b}^* = (1/4, 3/4, 0)$ when X_1 is true. According to the symmetric Brier rule, your initial inaccuracy is 1.10, and your final inaccuracy is 1.13, so your inaccuracy goes up. And this example works equally well against the symmetric spherical rule: according to this rule, your initial inaccuracy is 1.24 and your final inaccuracy is 1.37, so your inaccuracy goes up. This particular counterexample does not work against the symmetric log rule, but a similar one does. Suppose your initial credences are $\mathbf{b} = (1/13, 6/13, 6/13)$, and your final credences are $\mathbf{b}^* = (1/7, 6/7, 0)$. Then according to the symmetric log rule your initial inaccuracy is 3.80, and your final inaccuracy is 3.89: your inaccuracy goes up. Hence the symmetric measures appropriate to a Boolean algebra all violate Elimination.

However, the second consideration arguably suggests that the violation of Elimination is irrelevant as far as the proof of probabilism goes. The Elimination principle is a substantive diachronic assumption: as we argued in section 3, Elimination gains its most direct support from thinking about epistemic progress in science. A proof of probabilism can arguably proceed independently of any specific claims about learning and epistemic progress.[12] Hence we can arguably accept measures of inaccuracy that violate Elimination to vindicate probabilism. Note, however, that such a proof of probabilism is fragile: if we adopt the more substantive diachronic assumptions suitable for deriving conditionalization, including Elimination, then the symmetric measures are no

[11] Strictly, applying these rules to a Boolean algebra requires including credences in the negations $\neg X_1$, $\neg X_2$ and $\neg X_3$, plus the tautology $X_1 \vee X_2 \vee X_3$ and the contradiction $\neg(X_1 \vee X_2 \vee X_3)$. But for coherent credences the inaccuracies of the tautology and the contradiction are zero, and for coherent credences and symmetric rules the inaccuracy of $\neg X_i$ is the same as that of X_i. So the inaccuracy calculated over the entire Boolean algebra is simply twice the inaccuracy over the partition (X_1, X_2, X_3).

[12] Note that Blackwell and Drucker (2019) argue that *Strict Propriety* is a substantive epistemic assumption; certainly its typical justification is diachronic, as it appeals to the irrationality of credence shifts in the absence of new evidence (Maher 1990, 112). If Blackwell and Drucker are right, then probabilism cannot be proven without substantive epistemic assumptions. If we are right about Elimination, then probabilism cannot be proven *with* substantive epistemic assumptions.

longer acceptable, and probabilism is no longer provable. Furthermore, Elimination might also be interpreted as a *synchronic* constraint on credences, by taking $\mathbf{b}$ and $\mathbf{b}'$ to be two distinct sets of possible credences at a single time. If this synchronic version of Elimination is plausible, then probabilism cannot be proven *at all*, given the three main inaccuracy measures.

Let us sum up. The simple Brier rule cannot be used to prove conditionalization, but the simple log and spherical rules can. In the best-case scenario, where Elimination is taken to be a purely diachronic constraint on credence *shifts*, the symmetric Brier, log and spherical rules can be used to prove probabilism, but none of them also underwrites conditionalization. In the worst-case scenario, where Elimination is taken to be also applicable synchronically, none of the rules we have considered can be used to prove probabilism. Either way, we have found no measure that can be used to prove both conditionalization *and* probabilism.

5 Elimination defended

We have argued that none of the prominent measures of inaccuracy allow one to prove both probabilism and conditionalization. In discussion, the most frequent response to this argument is to deny Elimination—to deny that elimination cases are always epistemically positive. So let us spend some time defending this claim. The positive case for Elimination was laid out in section 3. What of arguments against it?

It is tempting to think that it doesn't matter whether a measure violates Elimination: as long as the measure is strictly proper, conditionalization is guaranteed to minimize *expected* inaccuracy, and since you cannot know whether you are reducing your *actual* inaccuracy, minimizing expected inaccuracy is all you can hope to achieve. However, even if, from a first-person point of view, elimination of a false hypothesis seems reasonable (because it minimizes the expected value of your chosen measure), it is a separate question whether, from a God's-eye point of view, you have actually made an accuracy improvement.[13] Our contention is that measures of inaccuracy that violate Elimination sometimes give the wrong answer to this God's-eye question.

The fact that measures that violate Elimination sometimes give the wrong answer to the God's-eye question shows that such measures don't measure *inaccuracy*, and hence that you have no reason to adopt a policy of minimizing the expected value of such measures. We do not dispute the Greaves and Wallace proof—conditionalization minimizes expected inaccuracy for any strictly proper *inaccuracy* measure. But not every strictly proper function of credences is a measure of *inaccuracy*. Consider, for example, the following weighted Brier rule:

Weighted Brier rule: $I(\omega, \mathbf{b}) = \sum_i \lambda_i \left(b_i - \omega_i\right)^2$,

[13] Fallis (2002, 227) and Dunn (2018, 9) stress this point.

where $\lambda_1 = 1$ and $\lambda_{i \neq 1} = 10^{-6}$. This weighted Brier score is strictly proper. But it entails that only your credence in X_1—say, the proposition that squirrels eat nuts— makes a significant contribution to your overall inaccuracy, your credences in all other propositions contributing only negligibly. This particular proper scoring rule is not a good candidate for an inaccuracy measure; it doesn't measure your actual inaccuracy.[14] So the fact that conditionalization minimizes the expected value of this measure is irrelevant to what you should do, epistemically speaking. Similarly, we argue that since the simple Brier rule and the symmetric Brier, log and spherical rules violate Elimination, they don't measure inaccuracy. Hence there is no epistemic reason to try to minimize their value, and they cannot ground conditionalization.

A different defense of measures that violate Elimination concerns prior credences. Note that cases in which the various rules violate Elimination are ones where your prior credences are highly inaccurate. Perhaps violations of Elimination are to be expected given inaccurate prior credences, even for perfectly good measures of inaccuracy. For example, one might think that conditionalizing from highly inaccurate priors increases your inaccuracy because your priors "aim" the credence shift resulting from conditionalization.[15] This is certainly true for the simple Brier rule, as our examples show: if Tom did it, and you conditionalize on the falsity of "Harry did it", then more accurate priors will result in a shift toward the truth, and less accurate priors will result in a shift away from the truth. But is there any independent reason to think that prior credences should affect the "aim" of conditionalization in this way? As noted in section 3, the evidence that Harry didn't do it is not, in itself, misleading—so it is hard to see why it should mislead. And to assume that inaccurate priors preclude making epistemic progress conflicts with assumptions about scientific epistemology, as also explained in section 3.

Perhaps the idea is that when your initial credence in the true hypothesis is sufficiently small, the large increase in credence in false hypotheses surely outweighs the small increase in credence in the true hypothesis. The Elimination principle entails that a shift from $(0.001, 0.4995, 0.4995)$ to $(0.002, 0.998, 0)$ when the first element is true is epistemically positive, but this might strike one as absurd given that your credence in the truth shifts by a tiny amount and your credence in a false hypothesis increases by almost 0.5. However, this description of the credence shifts assumes that the appropriate comparison is between the *absolute* values of the two credence shifts. Following Hacking (1965, 70) and Sober (2008, 32–34), we prefer to think of credence as a measure of *relative* confidence, in which case the relevant comparison is between the *ratios* by which the two credences have shifted.[16] In this case, both credences increase by exactly the same factor (almost 2). We contend that this is what it is for a shift to be epistemically neutral: this is what conditional-

[14] Dunn (2018) argues that certain weighted Brier rules are good candidates for a combined measure of inaccuracy and verisimilitude. But *this* measure is a hopeless measure of *inaccuracy alone*.

[15] We thank James Joyce for this suggestion.

[16] See also Fallis and Lewis (2016, 582).

ization produces when the evidence is entirely uninformative between the first two elements. Since the change in credence in the third element is clearly an epistemic improvement, the net result is epistemically positive.

Dunn (2018, 9) suggests that eliminating a false hypothesis might be misleading by way of an analogy with categorical belief. Suppose you believe a large number of false conditionals with the same true antecedent. Then learning that antecedent (i.e. eliminating its negation) can lead you to acquire a large number of false beliefs, and hence can be misleading. However, this is not a particularly close analogy to the cases we are interested in: we are interested in cases where you adjust your credences in a fixed range of propositions, whereas Dunn's case involves the production of new categorical beliefs. We could make the case more closely analogous, but then it is not clear that elimination of a false hypothesis would be misleading. If you have high credence in a large number of false conditionals with the same true antecedent, and you learn that antecedent, then typically your credence in those conditionals will go down.[17]

Oddie (2017, 19) and Dunn (2018, 10) argue that verisimilitude considerations show that Elimination is false. Their basic point is that elimination of a false hypothesis can result in credence being shifted from a false hypothesis that is close to the truth to a false hypothesis that is further from the truth, resulting in overall credences that are further from the truth. Dunn (2018, 10) provides a detailed counterexample to Elimination along these lines. Suppose you are interested in propositions A, B and C, which are not mutually exclusive or exhaustive, and hence do not form a partition. Nevertheless, we can form a partition out of appropriate compounds: $(ABC, AB\bar{C}, A\bar{B}C, A\bar{B}\bar{C}, \bar{A}BC, \bar{A}B\bar{C}, \bar{A}\bar{B}C, \bar{A}\bar{B}\bar{C})$. Suppose that ABC is true, and your initial credences in this partition are (0.0033, 0.0033, 0.0033, 0.19, 0.8, 0, 0, 0). If you conditionalize on the evidence that A is true, eliminating the false proposition $\bar{A}BC$, your new credences are (0.0167, 0.0167, 0.0167, 0.95, 0, 0, 0, 0). The first element in this partition is true, and the second, third and fifth are close to the truth, as they differ from the true element only in the truth value of one of the atomic components. Hence conditionalization primarily involves increasing credence in a proposition far from the truth $(A\bar{B}\bar{C})$ at the expense of credence in a proposition close to the truth $(\bar{A}BC)$. This, Dunn writes, is "a bad trade" (2018, 11).

We do not dispute that verisimilitude can trump accuracy in this way: we intend Elimination to apply only to accuracy measures, not to a combined measure of accuracy and verisimilitude (if such a thing is possible).[18] Furthermore,

[17] Suppose your initial credences in the partition $(AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B})$ are $(0.3, 0.1, 0.3, 0.3)$, so your credence in A is 0.4, your credence in B is 0.6, and your credence in $(A \to B)$, interpreted as a material conditional, is 0.9. If you learn A, your credences in the partition become $(0.75, 0.25, 0, 0)$, so your credence in A is 1, your credence in B is 0.75, and your credence in $(A \to B)$ is 0.75. This does not look misleading: although your credence in the false hypothesis B has increased a little, your credence in the true hypothesis A has increased a lot, and your credence in the false conditional $(A \to B)$ has decreased.

[18] Oddie (2017) argues against the possibility of a combined measure; Dunn (2018) and Schoenfield (2019) argue in favor. We remain agnostic.

it is sometimes hard to disentangle accuracy considerations from verisimilitude considerations in evaluating a particular case. Nevertheless, there are cases in which verisimilitude is beside the point, and in those cases it looks like Elimination holds. The Tom-Dick-Harry example is one such case. The three basic hypotheses are (presumably) equally far from each other in verisimilitude terms. If we include the rest of the Boolean algebra, the only other non-trivial elements are the negations (Tom didn't do it, Dick didn't do it, Harry didn't do it), which again are presumably equally far from each other. Here verisimilitude is irrelevant, and Elimination is vindicated.

Dunn (2018, 20) has a second criticism of Elimination, one that is not based on verisimilitude considerations. He points out that those elimination cases that we take as counterexamples to the simple Brier rule (and the other rules discussed in section 4) all have the feature that conditionalization *concentrates* your credence on fewer false hypotheses. Our counterexample to the simple Brier rule involves a shift from $(1/7, 3/7, 3/7)$ to $(1/4, 3/4, 0)$ where the first element is true; hence credence in the two false hypotheses is initially equally distributed, but most of it becomes concentrated on just one of them. It is typically (although not universally) acknowledged that spreading credence evenly over false hypotheses decreases inaccuracy, and concentrating it increases inaccuracy. Hence Dunn (2018, 21) contends that $(1/4, 3/4, 0)$ really is less accurate than $(1/7, 3/7, 3/7)$ when the first element is true, just as the simple Brier rule says, precisely because the credence in false hypotheses is more concentrated.

Although the relevance of falsity distribution to accuracy is sometimes contested, we do not wish to deny it here.[19] That is, we accept for the sake of argument that there are cases in which concentration of credence on fewer false hypotheses leads to an increase in inaccuracy, namely cases in which this concentration is not counterbalanced by any inaccuracy-decreasing shift.[20] For example, the shift from $(1/7, 3/7, 3/7)$ to $(1/7, 6/7, 0)$ when the first element is true clearly increases inaccuracy. But in the counterexample to the simple Brier rule, there is a counterbalancing factor, namely the increase in credence in the true hypothesis. The question, in such cases, is where to draw the line between shifts that increase overall inaccuracy and shifts that decrease it.

First, note that in order to vindicate the simple Brier rule, concentration of credence on fewer false hypotheses must carry a lot of epistemic weight. Consider a variant on Dunn's (2018, 10) proposed counterexample to Elimination: your initial credences in the partition $(ABC, AB\bar{C}, A\bar{B}C, A\bar{B}\bar{C}, \bar{A}BC, \bar{A}B\bar{C}, \bar{A}\bar{B}C, \bar{A}\bar{B}\bar{C})$ are $(1/6, 1/2, 1/6, 1/6, 0, 0, 0, 0)$, and you conditionalize on the information that C is false, resulting in credences $(0, 3/4, 0, 1/4, 0, 0, 0, 0)$. If $A\bar{B}\bar{C}$ is true, the Brier rule entails that your epistemic situation becomes

[19] See Knab and Schoenfield (2015) for an argument that falsity distributions are irrelevant to accuracy. Note that the simple log rule ignores falsity distributions: accuracy is a function only of credence in the true hypothesis.

[20] Cases in which your credence in the true hypothesis is zero are exceptions, as explained shortly.

worse.[21] But you have eliminated a false hypothesis, which would seem to make your epistemic situation better. Furthermore, your credence has shifted from a hypothesis far from the truth (ABC) and a hypothesis close to the truth ($A\bar{B}C$) to a hypothesis equally close to the truth ($AB\bar{C}$) and the truth ($A\bar{B}\bar{C}$), so verisimiltude considerations also seem to make your epistemic situation better. The only factor that might make your epistemic situation worse is the concentration of credence on the false hypothesis $AB\bar{C}$. While such concentration may well decrease accuracy, it is hard to see why it should be accorded so much epistemic weight that it would overcome both elimination and verisimilitude considerations.

In fact, we think there is a clear answer to the question of where to draw the line between shifts that increase inaccuracy and shifts that decrease it. For positive a and b, a credence shift from $(a/(a+2b), b/(a+2b), b/(a+2b))$ to $(a/(a+2b), 2b/(a+2b), 0)$ when the first element is true clearly increases inaccuracy, as it concentrates credence on one false hypothesis while leaving credence in the true hypothesis unchanged. But a credence shift from $(a/(a+2b), b/(a+2b), b/(a+2b))$ to $(a/(a+b), b/(a+b), 0)$ when the first element is true decreases inaccuracy, we claim, because it is an elimination case, and scientific epistemology demands that elimination cases decrease inaccuracy. At $a = 0$ these shifts coincide. Hence this must be the dividing line: the shift from $(0, 1/2, 1/2)$ to $(0, 1, 0)$ when the first element is true must be neutral regarding inaccuracy.[22]

Dunn (2018, 18) correctly notes that Elimination counts shifts like the latter as elimination cases, and hence as decreasing inaccuracy. Hence Elimination needs a slight modification:

Elimination*: For coherent credences over a partition, if $\mathbf{b}$ assigns a zero credence to some false proposition to which $\mathbf{b}'$ assigns a non-zero credence, and credences in the remaining propositions stay in the same ratios, then $\mathbf{b}$ is at least as accurate as $\mathbf{b}'$, and $\mathbf{b}$ is more accurate unless the credence in the true proposition is zero.

With this modification, Elimination* allows that concentrating credence on fewer false hypotheses contributes to inaccuracy, and appropriately delimits

[21] When the fourth element is true, the inaccuracy of $(1/6, 1/2, 1/6, 1/6, 0, 0, 0, 0)$ is 1, and the inaccuracy of $(0, 3/4, 0, 1/4, 0, 0, 0, 0)$ is 9/8. Dunn (2018, 15) defends a weighted Brier score to combine accuracy considerations with verisimilitude considerations; in cases like this he suggests that most of the weight should fall on the atomic statements (A, B, C). Initially, your credences in (A, B, C) are $(1, 2/3, 1/3)$, yielding a Brier score of $5/9 = 80/144$, and after conditionalizing they are $(1, 3/4, 0)$, yielding a Brier score of $9/16 = 81/144$. So whether you use a weighted or an unweighted Brier score, the score tells you that your epistemic situation has become worse.

[22] The simple spherical rule has this consequence. The consequence entails that the epistemic benefit of eliminating a false hypothesis scales with your credence in the truth, becoming zero when your credence in the truth is zero. This seems quite intuitive. It also entails that the epistemic cost of concentrating credence on fewer false hypotheses scales with your credence in the truth, becoming zero when your credence in the truth is zero. This is less intuitive, but seems unavoidable.

the cases where this effect results in an overall increase in inaccuracy. We conclude that there are strong reasons to accept Elimination*, and no compelling reason to abandon it.

6 The extent of the problem

If Elimination* is correct, then we have argued that none of the inaccuracy measures in the literature can be used to prove both conditionalization and probabilism. But could there be a measure from which both conditionalixation and probabilism can be proved that has not been described yet? We cannot rule that out. However, it is worth noting that *any* inaccuracy measure that satisfies Additivity, Strict Propriety and a plausible symmetry principle is subject to elimination counterexamples, as we show in this section. The symmetry in question is the one displayed by the symmetric Brier, log and spherical rules: the inaccuracy measure treats truth the same as falsity, in the sense that inaccuracy is always the same function of the distance between each credence and the truth value of the corresponding proposition, regardless of whether that proposition is true or false. For an additive inaccuracy measure, the symmetry principle can be expressed in terms of the inaccuracy function for a single proposition $s(\omega_i, b_i)$ as follows:[23]

Symmetry: $s(\omega_i, b_i) = s(|1-\omega_i|, |1-b_i|)$.

It is certainly highly plausible that this is part of what it means for s to measure the accuracy of your credences, and as discussed in section 1, the typical Boolean algebra forms of the Brier rule, log rule and spherical rule all satisfy it.

Let us see how Symmetry, together with Additivity and Strict Propriety, lead to elimination counterexamples. Consider a single proposition X_i in which your credence is $b_i = 1/2$. According to Strict Propriety, the quantity $\frac{1}{2}s(1,x) + \frac{1}{2}s(0,x)$ must be uniquely minimized at $x = 1/2$. In particular, the value of this expression for $x = 1/2$ must be lower than its value for $x = 1$:

$$\tfrac{1}{2}s(1,1/2) + \tfrac{1}{2}s(0,1/2) < \tfrac{1}{2}s(1,1) + \tfrac{1}{2}s(0,1),$$

and for $x = 0$:

$$\tfrac{1}{2}s(1,1/2) + \tfrac{1}{2}s(0,1/2) < \tfrac{1}{2}s(1,0) + \tfrac{1}{2}s(0,0).$$

Adding these:

$$s(1,1/2) + s(0,1/2) < \tfrac{1}{2}s(1,1) + \tfrac{1}{2}s(0,1) + \tfrac{1}{2}s(1,0) + \tfrac{1}{2}s(0,0).$$

But by Symmetry, $s(1,1/2) = s(0,1/2)$, $s(1,1) = s(0,0)$ and $s(0,1) = s(1,0)$. Substituting:

$$2s(0,1/2) < s(0,1) + s(0,0).$$

[23] Joyce (2009, 274) calls this principle "0/1-symmetry".

Now consider your credences in three exhaustive and mutually exclusive propositions $\mathbf{X} = (X_1, X_2, X_3)$. Consider in particular the credence shift from $\mathbf{a} = (0, 1/2, 1/2)$ to $\mathbf{b} = (0, 1, 0)$ for truth values $\omega = (1, 0, 0)$. By Additivity, $I(\omega, \mathbf{a}) = s(1, 0) + 2s(0, 1/2)$, and $I(\omega, \mathbf{b}) = s(1, 0) + s(0, 1) + s(0, 0)$. So since $2s(0, 1/2) < s(0, 1) + s(0, 0)$ it follows that $I(\omega, \mathbf{a}) < I(\omega, \mathbf{b})$: your inaccuracy goes up. But as argued in section 5, and codified in Elimination*, the shift from $\mathbf{a} = (0, 1/2, 1/2)$ to $\mathbf{b} = (0, 1, 0)$ should leave your inaccuracy unchanged. Furthermore, consider the credence assignments $\mathbf{a}' = (\delta/(2+\delta), 1/(2+\delta), 1/(2+\delta))$ and $\mathbf{b}' = (\delta/(1+\delta), 1/(1+\delta), 0)$. For small δ these are close to $\mathbf{a}$ and $\mathbf{b}$, and hence by the continuity clause of Additivity, the inaccuracy of $\mathbf{a}'$ remains lower than that of $\mathbf{b}'$. But the transition from $\mathbf{a}'$ to $\mathbf{b}'$ is an elimination case, and now your credence in the true proposition is non-zero, so according to Elimination* your inaccuracy should go down.

So elimination counterexamples afflict any inaccuracy measure that satisfies Additivity, Strict Propriety and Symmetry. That is, any symmetric measure that satisfies the assumptions of Predd et al.'s proof of probabilism violates Elimination*, and hence cannot be used to prove conditionalization. Symmetry is not a premise in the Predd argument, so it is possible that some asymmetric measure might allow the derivation of both probabilism and conditionalization. But we know of no such measure, and any such measure would face the objection that Symmetry, too, seems like a plausible constraint on inaccuracy measures.

7 Conclusion

Pettigrew notes that conditionalization and probabilism follow from a wide range of measures of inaccuracy, and the implication is that it doesn't much matter which measure you pick. We have argued that, in the best-case scenario in which Elimination* is taken to be a purely diachronic constraint on credence shifts, there are measures that vindicate conditionalization, and there are measures that vindicate probabilism, but nobody has yet identified a measure that vindicates both. In the worst-case scenario, in which Elimination* is also applicable synchronically, none of the available measures can be used to prove probabilism. Hence the accuracy-based approach does not, as yet, give us the justification we might want for the two standard constraints on our credences.[24]

References

Blackwell, Kevin, and Daniel Drucker (2019), "When propriety is improper", *Philosophical Studies* 176: 367–386.

[24] We would like to thank Jeffrey Dunn, Kenny Easwaran, James Joyce, Brian Knab, Graham Oddie, Richard Pettigrew, Terry Horgan, the audience at the Philosophy of Science Association Biennial Meeting in Atlanta in November 2016, and two anonymous referees for very helpful comments on earlier versions of this paper.

Carr, Jennifer Rose (2017), "Epistemic utility theory and the aim of belief", *Philosophy and Phenomenological Research* 95: 511–534.

Dunn, Jeffrey (2019), "Accuracy, verisimilitude, and scoring rules", *Australasian Journal of Philosophy* 97: 151–166.

Earman, John (1992), *Bayes or Bust.* Cambridge, MA: MIT Press.

Fallis, Don (2002), "Goldman on Probabilistic Inference", *Philosophical Studies* 109: 223–240.

Fallis, Don and Lewis, Peter J. (2016), "The brier rule is not a good measure of epistemic utility (and other useful facts about epistemic betterness)", *Australasian Journal of Philosophy* 94: 576–590.

Greaves, Hilary and David Wallace (2006), "Justifying conditionalization: conditionalization maximizes expected epistemic utility", *Mind* 115: 607–632.

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

Hedden, Brian (2013), "Incoherence without exploitability", *Noûs* 47: 482–495.

Joyce, James M.. (1998), "A nonpragmatic vindication of probabilism", *Philosophy of Science* 65: 575–603.

Joyce, James M. (2009), "Accuracy and coherence: prospects for an alethic epistemology of partial belief", in F. Huber and C. Schmidt-Petri (eds.), *Degrees of Belief.* Dordrecht: Springer: 263–297.

Knab, Brian and Miriam Schoenfield (2015), "A strange thing about the Brier score", *M-Phi* blog, March 12, 2015. http://m-phi.blogspot.nl/2015/03/a-strange-thing-about-brier-score.html

Leitgeb, Hannes, and Richard Pettigrew (2010), "An objective justification of Bayesianism I: measuring inaccuracy", *Philosophy of Science* 77: 201–35.

Maher, Patrick (1990), "Why scientists gather evidence", *British Journal for the Philosophy of Science* 41: 103–119.

Marxen, Chad, and Gerard Rothfus (2018), "Review of Richard Pettigrew, *Accuracy and the Laws of Credence*", *Philosophy of Science* 85: 316–320.

Mayo, Deborah G. (1996), *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

Oddie, Graham (2017), "What accuracy could not be", *British Journal for the Philosophy of Science* 70: 551–580.

Pettigrew, Richard (2013), "Epistemic utility and norms for credence", *Philosophy Compass* 8: 897–908.

Pettigrew, Richard (2016), *Accuracy and the Laws of Credence.* Oxford: Oxford University Press.

Predd, Joel B., Robert Seiringer, Elliott H. Lieb, Daniel N. Osherson, H. Vincent Poor, and Sanjeev R. Kulkarni (2009), "Probabilistic coherence and proper scoring rules", *IEEE Transactions on Information Theory* 55: 4786–4792.

Popper, Karl R. (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge.* London: Routledge.

Schoenfield, Miriam (2019), “Accuracy and verisimilitude: The good, the bad and the ugly”, manuscript.
http://www.miriamschoenfield.com/F/propriety-25.pdf

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

Talbot, Brian (2017), “Repugnant accuracy”, *Noûs*, forthcoming.
https://doi.org/10.1111/nous.12237

Vineberg, Susan (2012), “Dutch book arguments”, in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy.*
http://plato.stanford.edu/archives/spr2012/entries/dutch-book/