Moral Obligations and Epistemic Risk

Boris Babic (Caltech) and Zoë Johnson King (NYU)

(draft for WiNE 2019, last updated Dec 11 2018)

1. Introduction

Tamar Gendler (2011, p.57, emphasis added) argues for what she calls "The Sad Conclusion", as follows:

> As long as there's a differential crime rate between racial groups, a perfectly rational decision maker will manifest different behaviors, explicit and implicit, toward members of different races. This is a profound cost: *living in a society structured by race appears to make it impossible to be both rational and equitable*.

The idea is illustrated in terms of racial injustice, but The Sad Conclusion is more general in spirit. It is informed by the general observation that, in societies structured by group-based injustices, there will be many statistical facts that are accurate but uncomfortable to think about, and perhaps morally and politically problematic to act on.

For example, suppose that you are newly hired by an investment bank. You are told that when you come to work, a partner of the firm and their associate will take you to lunch. Soon after you arrive, a male and female colleague come by your office. You are aware of the gender gap in financial institutions like the one that now employs you: in such firms, the overwhelming majority of partners are male. As a result, your initial guess is that the male colleague is likely to be the partner. Assuming that your prior information about base rates is accurate, this seems epistemically rational. But it also seems morally objectionable, since it amounts to an assumption about your female colleague based on information about a social group to which she belongs. Moreover, in your context, this is a *pernicious* assumption about your female colleague. You assume that the male colleague has the property that is socially valued in this context — being a partner — and that the female colleague has the less-valued property of being an associate.

There are many reasons why this assumption seems morally and politically problematic. It contributes to a situation in which being female is associated with subordinate positions in the workplace, while being male is associated with positions of higher authority. These associations can lead, and have led, to an inhospitable environment for women in the workplace. In other cases — where base rates associate criminality with race or ethnicity, for example — these associations can lead, and have led, to serious harms for members of socially marginalized groups. They can also be self-reinforcing, which produces a multiplier on the harms experienced by the relevant groups that goes beyond the underlying base rates. In addition, some philosophers (e.g. Basu and Schroeder 2018) have argued that it is possible to wrong somebody simply by believing pernicious things about them. If this is true, then presumably it is possible to wrong someone by *degrees* corresponding to one's degree of belief in the pernicious thing, and this presumably holds of your expecting your female colleague to be the more junior one.

The idea behind the Sad Conclusion is that, notwithstanding all of these moral and political difficulties, epistemic rationality requires you to to expect your female colleague to be the associate rather than the partner. For epistemic rationality is concerned solely with the accuracy of our doxastic states, and is independent of moral and political considerations. The Sad Conclusion is that, if it wrongs people to assume pernicious things about them based on statistical information about groups of which they are members, then that's just too bad — such assumptions are frequently epistemically required, and so wronging people is frequently epistemically required. This way of thinking portrays epistemic rationality as a harsh taskmaster, coldly insisting that our doxastic states must "respect the data" and be indifferent to the moral and pragmatic context within which they operate.

This paper brings good news: the Sad Conclusion is false. Requirements of epistemic rationality do not inevitably conflict with those of morality in societies structured by group-based injustices. Moreover, we need not adopt a conception of epistemic rationality that is unfriendly toward data and probabilistic reasoning in order to secure this result. On the contrary, one of the most technical and statistician-friendly views of the requirements of epistemic rationality entails that *no* particular attitude is required by epistemic rationality in cases like the one described above. This happier conclusion follows from a version of *accuracy-based epistemology* according to which epistemic agents are rationally required to form and revise their *credences* so as to *maximize expected epistemic utility*, where epistemic utility is understood in terms of the accuracy of the agent's credences and measured by a kind of function known as a *scoring rule*, where the function's shape depends on the agent's attitude toward *epistemic risk*. Or so we argue in this paper. We go over the basics of this way of thinking about the requirements of epistemic rationality in §2b; terms in italics in this introduction are explained later in the paper.

On our preferred approach, epistemic rationality permits a wide range of attitudes to epistemic risk. Rationality requires that we form and revise credences with the aim of believing truths and disbelieving falsehoods, in a way that satisfies certain formal symmetry and coherence constraints that collectively ensure that we pursue these aims in a logically consistent manner. But these formal constraints do not settle the question of how to assess the relative badness of a high credence in a falsehood and a low credence in a truth: the question of how to assess the *relative costs of different types of error.* Many ways to assess the relative costs of different types of error respect epistemic rationality's formal constraints, as is well-known in the extant literature on accuracy-based epistemology.

Moreover, we hold that there is nothing amiss from an epistemic point of view in an agent's taking the moral and political costs of different types of error into account in her overall assessment of these costs. Indeed, we hold that one cannot avoid taking some attitude or other toward these costs. Formal epistemologists sometimes say that when an agent does not have any information about a set of outcomes, epistemic rationality requires her to assign the same probability to each outcome. But we will argue that, rather than remaining neutral on the moral and political costs of error, such a principle adopts a very specific attitude to epistemic risk: the attitude that errors with respect to any of these possible outcomes are all exactly equally as bad as one another.

More generally, an agent's epistemic risk profile reflects her normative commitments, where these commitments are expressed in the way she values assigning high credences to falsehoods and assigning low credences to truths. These attitudes to epistemic risk shape her *measure* of inaccuracy — that is, her way of evaluating inaccuracy. This means that the agent's attitudes toward the moral and political impact of disbelieving a truth or believing a falsehood partly determine which credences she should adopt in the sorts of contexts thought to give rise to the Sad Conclusion. It follows that the Sad Conclusion is false: on a plausible view about what rationality requires, if morality requires us to adopt a certain attitude in cases like the one described above, epistemic rationality typically permits this attitude.

The above applies to *predictive inferences*: inferences about how likely it is that an encountered individual possesses a certain property, based on data about the prevalence of the property within a sample of a social group to which she belongs. This is the sort of case usually raised as morally problematic in the existing literature on this topic (as we will see in §2.1). But we argue that matters are morally and epistemically different when it comes to statistics about *populations* – for example, the statistic that X% of executives in Major League Baseball are female. This is a case where we know the distribution of a characteristic in an entire population, rather than estimating what the population distribution might be based on a sample from it. In these cases, our framework entails that epistemic rationality does have something to say: we should accept statistical information about population frequencies. But, we argue (in §4), this is not morally problematic. On the contrary, it is crucial to acknowledge accurate-but-uncomfortable statistical facts about populations, as it would be impossible to recognize and begin to address the underlying structural social injustices without doing so. So the Sad Conclusion is false again; when epistemic rationality does require us to adopt a certain attitude, morality typically permits it.

In §5 we discuss a final variety of statistical inference that complicates matters. The Sad Conclusion may hold true in certain cases of what is known as *direct inference*, as opposed to predictive inference: direct inference involves making a prediction about an individual based on data about a population sample, where the individual herself was included in the sample that gives rise to the data. (Reasoning from population statistics to predictions about members of the population is a type of direct inference, with the "sample" here being the entire population.) Formal coherence constraints apply differently to predictive and direct inference, and the differences matter for our purposes. In direct inference, mathematical coherence requires us to conform our credences to known population frequencies when we are under conditions of *exchangeability*: roughly, when we have no additional evidence making it more or less likely that the individual under discussion bears the property in question than any other individual from the sample. This means that it is conceivable that the sort of normative conflict that troubles Gendler could arise, since it arises for direct inference under conditions of exchangeability. Fortunately, though, such cases are vanishingly rare. The normal cases that have worried philosophers writing on this topic are cases in which someone makes an inference about a particular, identified individual that they have encountered. In these cases, the other evidence that the agent acquires about the person on encountering them constitutes an *informative label*, which means that she is no longer under conditions of exchangeability. Once again, then, the Sad Conclusion fails to hold.

2. Background: Statistical evidence and accuracy-based epistemology

a. *Statistical evidence*

Sometimes, statistical evidence suggests that a trait that is socially disvalued is more prevalent in some social groups than others. For example, statistical evidence might suggest that certain social groups tend to occupy positions of prestige in the workplace at lower rates than others (as in our example above), or to have higher rates of arrest and imprisonment than others, or to have higher rates of certain socially disvalued medical conditions than others. Call such socially disvalued traits "pernicious" traits, and call the inference that an individual possesses a pernicious trait based on statistical evidence about its prevalence in a social group to which she belongs a "pernicious predictive inference". Most people have the intuition that there is something morally problematic about pernicious predictive inference. This problem is widely discussed in legal contexts, in which it is usually thought impermissible to convict a defendant based on statistical evidence about crime rates in social groups to which she belongs.[1] More recently, the issue has received much discussion among philosophers interested in exploring the relationships between epistemic and moral norms.

The existing literature tends to focus on the negative. Philosophers consider the person who draws a pernicious predictive inference — for example, someone who assumes that her female colleague is more junior than her male colleague in the case described above — and argue (or have the intuition) that there is something morally defective with this inference. To avoid the Sad Conclusion, these philosophers look for something epistemically defective in the inference. The possibilities here are many and varied. The pernicious predictive inference may exhibit a general epistemic flaw; perhaps the agent over-extends the application of a statistic beyond the population from which it was

[1] For example, Tribe (1971) casts doubt on the use of 'naked statistics' in legal trials. Thomson (1986) suggests that statistical evidence may be inappropriate because it is not properly connected to the individual defendant, and Buchak (2014) suggests that while statistical evidence can support a high credence in a defendant's guilt, it cannot support a belief which, Buchak argues, is a more appropriate doxastic attitude for blame and liability.

derived (see Munton *ms*), thereby failing to display understanding of the causal mechanisms that give rise to the pattern that it reports (see Gardiner *ms*). Or perhaps she goes beyond what her evidence supports in assuming that the female colleague is more junior, when her evidence only supports the proposition that the female colleague is *probably* more junior (*ibid.*, p.** — though we think that a belief that the female colleague is probably more junior is only marginally less objectionable than an outright belief that she is more junior, and so we think that this is not a good explanation). Alternatively, the pernicious content of the belief might itself be relevant to its epistemic status, such that pernicious predictive inference exhibits a kind of epistemic flaw that would not be exhibited by an equally statistically-well-supported belief with morally neutral content. Perhaps the fact that the trait is pernicious raises the amount of evidence required for epistemic justification to a greater amount than that which the subject possesses (see Basu *ms*). Or perhaps the fact that the trait is pernicious means that the subject must do more to rule out relevant alternatives than she has done (see Moss 2016).

These are all interesting and promising proposals, and we have no intention of arguing against them here. But it is noteworthy that all of these existing proposals focus on finding an epistemic fault with the pernicious predictive inference, in the context of an intuition or argument to the effect that such an inference is morally prohibited.

We focus on the positive. In this paper, we are interested in defending the epistemic credentials of the sort of predictive inference that seems morally praiseworthy, and perhaps even morally required: a predictive inference that exhibits reluctance to believe something bad about a particular individual based on demographic information about the prevalence of a pernicious trait within a social group to which she belongs. We take it for granted that such reluctance is morally good, and we are interested in exploring its epistemic status. This is an important shift of focus. Notice that, for the Sad Conclusion to be false, it does not need to be the case that the morally prohibited pernicious predictive inference is epistemically prohibited. Instead, we can show that the Sad Conclusion is false by showing that the morally good inference is epistemically permitted — which is the case if, as we think, epistemic rationality is silent on the matter. If this is so, then, *pace* Gendler, the requirements of morality do not inevitably conflict with those of epistemic rationality. On the contrary, there is a doxastic option on which morality smiles and at which rationality shrugs. By vindicating the epistemic status of the morally good inference, we secure the possibility of an all-things-considered normatively attractive option in cases of predictive inference concerning pernicious traits. Moreover, we secure this possibility without holding that the requirements of morality "trump" those of epistemic rationality in determining what someone all-things-considered ought to believe. By analogy: if Boris wants to go to Tomukun for lunch, while Zoë has no preference, and we decide to go to Tomokun, Boris has not "trumped" Zoë. Similarly, if epistemic rationality permits a wide range of attitudes toward epistemic risk, and morality favors some of these over others, to hold that the agent all-things-considered ought to adopt one of the attitudes that morality favors is not to hold that morality "trumps" or "beats" or "wins against" epistemic rationality. On the contrary, in such a case it is a mistake to think that the two are "against" one another at all.

There is a practical reason why we want to resist the idea that the morally praiseworthy predictive inference in these cases is epistemically inferior to the pernicious predictive inference. In our present social context, the suggestion that morally praiseworthy attitudes are in some way epistemically flawed can be used as a smokescreen behind which prejudicial attitudes hide. Prejudiced people sometimes try to convey the impression that cold, calculating rationality is on their side, and that reluctance to infer pernicious things about individuals based on statistical information about social groups to which they belong is a sign of weakness or a denial of the "hard facts". We think that this suggestion is worth challenging in and of itself, irrespective of the merits of the aforementioned philosophical projects developing epistemic criticisms of pernicious predictive inferences. We challenge the suggestion in this paper. We argue that epistemic rationality permits morally praiseworthy inferences — and, further, that this is a straightforward consequence of the most widely adopted formal way of thinking about what epistemic rationality requires, informed by contemporary work in statistics and probability theory. In short: people who hold prejudicial attitudes can't claim that rationality or mathematics are on their side, because, when properly understood, it is easy to see that they aren't. We take this dialectical point to be one of the main contributions of our paper.

b. *Accuracy-based epistemology*

In this paper we take an accuracy-based approach to epistemology. The basic ideas here are that the attitude of belief comes in degrees, that a higher degree of belief is more accurate than a lower one if the proposition in question is true, while a lower degree of belief is more accurate than a higher one if the proposition is false, and that it is better for one's degrees of belief to be more accurate than less accurate. In accuracy-based epistemology, it is standard to model degrees of belief – known as "credences" – using real numbers between 0 and 1, where 0 represents certainty that a proposition is false and 1 represents certainty that it is true. So, if a proposition is true then credence 0.8 is more accurate than credence 0.2, whereas if it is false then credence 0.2 is more accurate than credence 0.8.

It is an open question exactly how accuracy and inaccuracy should be measured. A foundational assumption of accuracy-based epistemology is that higher credences in true propositions are more accurate than lower ones, while lower credences in false propositions are more accurate than higher ones. Formally, this means that inaccuracy is a *truth-directed monotonic function* of the agent's credence in the relevant proposition. It is also natural to think that our measure of inaccuracy should be *continuous* – i.e., that arbitrarily small changes in degrees of belief should not result in big leaps in inaccuracy. In keeping with these assumptions, inaccuracy is typically depicted as a monotonic and continuous two-place function s: $[0, 1] \text{ x } \{0, 1\} \rightarrow \mathbb{R}$ taking as inputs the agent's credence and the actual truth-value of a proposition ('0' if false, '1' if true) and returning a real number representing her degree of inaccuracy. This function is known as a "scoring rule", since it "scores" the agent's credences for inaccuracy.

Most formal epistemologists think that, as well as being truth-directed and continuous, a good scoring rule must be *strictly proper*. This means that, when an agent uses a scoring rule to evaluate the accuracy of various different sets of credences, her own current credences always look best in expectation: they have the lowest expected inaccuracy. An improper scoring rule can leave an agent in a hopeless grass-is-greener situation, such that, no matter what her credences are, some other set of credences looks better to her — but if she were to switch to this set of credences, then another set of credences would look better to her. Following an improper scoring rule would thus prevent our credences from ever being stable. Most formal epistemologists reject improper scoring rules on these grounds. It is possible for a scoring rule to be *proper*, but not strictly proper: this is so when the scoring rule evaluates some other possible sets of credences as equal in expected inaccuracy to the agent's current credences, but none as doing better. Following such a scoring rule would permit the agent to stick to her current credences, unlike the improper scoring rules. Still, a proper-but-not-strictly-proper scoring rule would also permit her to switch to one of the equally-good sets of credences, which amounts to suddenly and arbitrarily shifting to a totally different set of credences without any change in one's evidential state. This also seems inappropriate. Thus, formal epistemologists typically assume that scoring rules should be (1) monotonic, (2) continuous, and (3) strictly proper. We refer to these three properties collectively as the "accuracy checklist". A scoring rule that satisfies the accuracy checklist is a good scoring rule.

It is natural to think that credences in some propositions rationally commit an agent to having certain other credences in related propositions. For example, suppose that Zoë has credence 0.7 that Boris is at the gym. Surely, Zoë should have credence 0.3 that Boris is not at the gym. This is because those two propositions are mutually exclusive and jointly exhaustive of logical space; either Boris is at the gym or he isn't, and he can't be both there and not there. A set of propositions that are mutually exclusive and jointly exhaustive of logical space like this is called a *partition*. It is natural to think that one's credence in a disjunction of the elements of a partition should be equal to the sum of one's credences assigned to each disjunct. This (finite) additivity assumption is one of a set of minimally strong axioms for credences that collectively ensure that an agent's credences obey the basic laws of probability. The second is that the credence assigned to any given proposition must be greater than or equal to 0.

Finally, we typically assume that the sum of an agent's credences in all elements of a partition must be 1, so as to have a common scale for comparing degrees of belief. We call someone *coherent* if her credences obey these three axioms, and *incoherent* if they don't. It is a central tenet of accuracy-based epistemology that epistemic rationality requires coherence: an agent's credences must obey the probability axioms. As well as being intuitively appealing, this tenet is supported by a well-known proof from James Joyce (1998, 2009) showing that, for any incoherent credence-function (i.e. set of credences), there is a coherent credence-function that is more accurate no matter how the world turns out to be, according to any scoring rule that satisfies the accuracy checklist. In short, any plausible way of evaluating accuracy entails that any incoherent credence-function is "accuracy dominated" by a coherent credence-function, which is to say that the latter is more accurate no matter what. Hence the idea that, for epistemic agents aiming at accuracy, coherence is rationally required.

When an agent gains new evidence relevant to a proposition, she should change her credences. For example, if you learn that a fair die will be tossed, then it is sensible to initially assume that the probability of it landing on 6 is 1/6. If you are then told that it landed on an even number, your credence that it landed on 6 should increase to 1/3. The credence that a die landed on a 6, given that it landed on an even number, is called a *conditional credence*. It is very natural to think if someone learns a new proposition (e.g., the die landed on an even number) then they should shift their prior credence in the proposition (e.g., the original credence that the die landed on 6, namely 1/6) to their prior *conditional* credence in the proposition, given the event they learned (e.g., the prior conditional credence that the die landed on 6, given that it landed on an even number, namely 1/3). Speaking somewhat informally, this is what it is to update by *Bayesian conditionalization*: in Bayesian updating the posterior credence in a proposition after learning that some event occurred is equal to the prior conditional credence in the proposition given the event one learns. The conditional credence for an event A, given an event B, is defined as P(A|B) = P(A & B)/P(B). In the die example, this gives us the result that the conditional credence in the die landing on a 6, given that it landed on an even number, should be (1/6)/(1/2) = 1/3.[2]

Just as Joyce vindicates coherence by showing that every incoherent credence function is accuracy dominated, given any scoring rule that satisfies the accuracy checklist, so too Greaves and Wallace (2006) show that updating by Bayesian conditionalization minimizes the prior expected inaccuracy of an agent's posterior credences, given any scoring rule that satisfies the accuracy checklist. This means that the best strategy for an agent who is deciding how to respond to some new evidence with the ultimate epistemic goal of minimizing her expected inaccuracy is to update by Bayesian conditionalization.[3]

We assume that epistemic rationality requires agents to assess accuracy using a scoring rule that is truth-directed, continuous, and strictly proper, to seek to minimize expected inaccuracy, and to hold coherent credences which they update by conditionalization. This can be fairly described as the generally accepted "core" of epistemic rationality concerning the formation and revision of credences. However, as we will explain, these requirements do not support the Sad Conclusion: they do not force us to adopt morally and politically problematic credences about individuals based on statistical data regarding the prevalence of pernicious traits among groups to which they belong. We will see why this is so in the next section.

3. Attitudes toward epistemic risk

[2] Note that the proposition that the die landed on a 6 and an even number is equivalent to the proposition that the die landed on a six, conditional on it landing on an even number, multiplied by the probability that it landed on an even number. That is, P(A & B) is equivalent to P(B|A)P(A). If we replace P(A & B) in the above definition of a conditional credence with P(B|A)P(A), then we get the usual statement of Bayes' Rule. This is why updating in this manner is known as Bayesian conditionalization.

[3] Leitgeb and Pettigrew (2010) and Easwaran (2013) provide similar results with subtle differences that are not relevant to this project.

a. *Rejecting the Frequency-Credence Connection*

Consider the following example:

> **Gender Bias Study**. One morning you read a report in the Washington Post about a study reporting gender discrepancies in academic employment in the United States. The authors of the study surveyed 1,000 people, 500 men and 500 women. They found that 70% of women held administrative roles and 30% held faculty roles, whereas the opposite was true of men. The study seems to have been conducted competently; the authors stipulated in advance that the survey would end after 500 men and 500 women responded, they did not perform multiple comparisons of their data in order to find evidence of gender bias, they disclosed all covariates that were tracked, and no observations were excluded. Before reading this article, you had no information about the relative distribution of women and men across different occupational roles in academia. After reading about the study you meet Mary, a new neighbor who tells you that she is employed at your local university, but does not tell you in what capacity she is employed.

What should your credence be that Mary is a faculty member? One natural-seeming answer is that you should assign credence 0.3 to the proposition that she is faculty, and credence 0.7 to the proposition that she is an administrative assistant. After all, you have no information regarding the proportion of women employed at universities who are faculty besides the results of the study. And the study's design raises no obvious red flags. So it is natural to think that, in the absence of any other relevant information, your estimate of the overall proportion of women employed at universities who are faculty should simply be equal to the proportion reported in the study. And it is natural to think that, in the absence of any other relevant information about Mary besides her gender and the fact that she is employed at a university, your credence that she is a faculty member should simply be equal to your estimate of the proportion of women employed at universities who are faculty — that is to say, the proportion reported in the study.

This natural answer is precisely the sort of answer that seems morally and politically troublesome. The assumption that Mary is not a faculty member raises the sorts of concerns we discussed in §2.1: it directly wrongs Mary, on Basu and Schroeder's (2018) account, and acting on it may lead you to treat Mary in a way that further wrongs her (e.g. by disparaging her in a way that reinforces negative social stereotypes about women in academia). This is the sort of case in which Gendler deems it impossible to be both rational and equitable. The information made available to you by this study forces you to make a pernicious predictive inference about Mary — or so the thought goes.

This is a mistake. The natural-seeming answer makes two errors about what epistemic rationality requires, and we can see that there is no conflict between the requirements of morality and those of epistemic rationality once the latter is stripped of these errors. The natural-seeming answer is wrong to think that epistemic rationality requires that your estimate of the overall proportion of women employed at universities who are faculty must be exactly equal to the proportion reported in the study (we discuss this mistake in this section). Indeed, doing so reveals an attitude that is naive in its responsiveness to evidence. And the natural-seeming answer is also wrong to think that, without other information about Mary besides her gender and the fact that she is employed at a university, epistemic rationality requires that your credence that she is faculty must be your estimate of the overall proportion of women employed at universities who are faculty. This is only true when certain conditions are met, which we will explain in more detail below (§5).

These purported rational requirements do not follow from the central tenets of accuracy-based epistemic rationality that we canvassed above. Indeed, the reverse is true; the predominant and most general version of the formal approach to thinking about epistemic rationality entails that these are *not* requirements. Thus, it follows from this view about epistemic rationality that the sort of reaction to **Gender Bias Study** that seems morally praiseworthy — a reluctance to make assumptions about Mary on the basis of the data in the study — is epistemically permissible.

To explain all of this, we have to unpack the natural-seeming answer in more detail, and to clarify some points on which the version just stated is a little rough.

The proposal is based on this principle:

> **Frequency-Credence Connection**: If (a) I know that *a* is an F, and (b) I know that *x*% of sampled Fs are G, and (c) I have no further evidence bearing on whether *a* is G, then my credence that *a* is G should be *x*.

The frequency-credence connection has a rich history. This version is from White (2010), but the principle is much older, and features in certain influential interpretations of early probability theory. For example, a version of it can be found in Reichenbach (1938, 1949), it was forcefully defended by Kyburg (1974), and it was further taken up in Levi (1977) — though Levi puts some important constraints on the principle's scope of application, which we will note in §5. Much recent work in formal epistemology continues to assume that the frequency-credence connection is a requirement of epistemic rationality, without paying too much attention to Levi's constraints, and some authors defend the Sad Conclusion on this basis (see e.g. Buchak 2014).

The intuitive idea behind the principle is that, when estimating the probability that a certain individual possesses a certain trait, we should conform our credences to known *population frequencies* — that is, facts about the frequency with which the trait has been observed to occur in samples of populations of which the individual is a member. (In **Gender Bias Study**, the trait is being a faculty member, and the population is the set of women who are employed at universities.) Notice that, in this kind of case, we are reasoning about an individual who was not part of the population used to determine the sample. Instead, we have a set of data about the women interviewed in the study, and we are using it to make a prediction about a new and different woman. This is called *predictive inference*.

Predictive inference is a little more complicated than the quick gloss on the natural-seeming answer that we gave above makes it sound. For illustration, consider the formal epistemologist's favorite example: a coin of an unknown bias. Suppose that Zoë is going to flip a coin, and Boris must guess whether the coin will land heads. Boris knows that the coin's weight distribution determines the probability that it will land heads on any given toss, so, in order to make a prediction about whether it will land heads, he must first estimate the coin's *bias* — its weight distribution. It might be a fair coin, such that its probability of coming up heads is 50%. Or it might be biased so that it comes up heads 90% of the time, or only 20%, or whatever. If he had already observed some tosses of the coin, then Boris could use these prior observations to guess the coin's weight distribution. But suppose that Boris has never seen any tosses of this coin before, and has no other information regarding its bias. Still, he knows that it is the weight distribution that determines the probability of the coin's landing heads. Moreover, Boris knows that the coin must have some particular weight distribution (i.e., it must be 99% likely to come up heads *or* 98% likely to come up heads *or*...), and it cannot have more than one weight distribution.

To make things simple, suppose that we "round" the weight distributions to the nearest tenth — so that the possible hypotheses for us are that the coin is 10% likely to come up heads, 20% likely to come up heads, and so on up until 100%. Then the set of hypotheses about the coin's weight distribution form a partition of logical space. Boris can assign credences to each of these hypotheses about the coin's weight distribution, and, to accord with the probability axioms stated in §2.2, these credences should sum to 1.

In certain important respects, **Gender Bias Study** is analogous to the case of the biased coin. There is an unknown quantity: the true proportion of women employed at universities who are faculty, rounded (for the sake of example) to the nearest tenth. It could be 50%, or 90%, or only 20%, or whatever. This proportion is analogous to the true weight distribution of the coin. And, just as the set of hypotheses about the coin's true weight distribution forms a partition of logical space, so does the set of hypotheses about the proportion of faculty among women employed at universities: there must actually be some such proportion, and there cannot be more than one. So, one can assign credences to each hypothesis about the true proportion, and these credences should sum to 1. For example, you might think it 20% likely that 60% of women employed at universities are faculty, 50% likely that 40% are faculty, and 30% likely that only 10% are faculty. But this distribution of your credence would be quite unusual. For a set of hypotheses that can be plotted along a continuum, like a range of proportions from 0% to 100%, people do not typically divide all their credence between just three hypotheses. In this case, people do not typically think that there is some chance of the true proportion of women faculty being 60%, some chance of its being 40%, and some chance of its being 10%, but no chance whatsoever of its being 50% or 20% or any other proportion. In other words, to model a person's credences about this unknown proportion we must now relax the "rounding" assumption, and assume that they assign some probability to any real number between 0 and 1 corresponding to the true proportion. Thus, their credences form a continuous distribution across the space of possible values, like so:

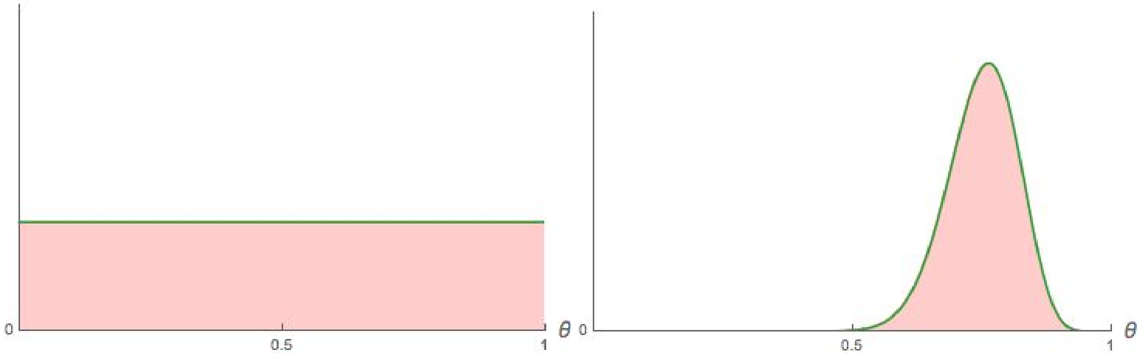

Fig. 1. *Two possible credal distributions.* ***θ*** *denotes the unknown proportion or bias, which can be any real value between 0 and 1. For any interval on the x-axis, the area shaded in pink is the relative density assigned to* ***θ*** *lying in this interval. For reasons that we omit here, the scale on the y-axis is not relevant. What matters is the relative mass assigned to any given interval — this mass ultimately corresponds to a measure of the probability that the true value of* ***θ*** *lies in that interval.*

This underlying space is no longer a partition. It is described by a set-theoretic concept known as a *sigma-algebra*: a collection of sets that respects basic set operations but is smaller than the set of all subsets on the space. We need not go into the details here. Rather, we will just say, informally, that one can think about the set of all the sub-intervals on the space of possible values from 0 to 1 as the appropriate analog to a partition.

The initial distribution of someone's credence over hypotheses about the value of an unknown quantity (e.g. the coin's bias, or the proportion of faculty among women employed at universities), prior to making any observations, is called a *prior distribution*. The figure above illustrates two possible prior distributions. The one on the left is *uniform*: it takes all hypotheses about the value of the unknown quantity to be equally likely. The one on the right is *non-uniform*: it regards some hypotheses about the value of the unknown quantity as more likely than others.

Now we can ask: what kind of information is sufficient to characterize the shape of these prior distributions? We could just write out the functions mathematically. For example, the distribution on the right is approximately bell-shaped (though not quite — the tails are shorter and that it is not symmetric around its mean), which can be described by the equation generating the graph. But we can do something simpler. To accurately describe the shape of the prior distribution in this kind of case, a theorem from Bayesian statistical theory states that it is sufficient to know two values, which can be interpreted as a set of imagined prior "favorable" and "unfavorable" observations — e.g., observations of heads and tails in prior tosses, or encounters with women employed at universities who are faculty members and those who are administrative assistants. We can use this heuristic of imagined favorable and unfavorable observations to characterize the relevant distributions. For example, in the left-panel of Figure 1, above, the plot corresponds to 1 imagined observation of heads/faculty and 1 imagined observation of tails/admin, whereas the right hand side corresponds to 30 imagined observations of heads/faculty and 10 of tails/admin.[4]

When an agent then makes a series of observations, she updates her estimates of the true value of the unknown quantity (the coin's bias, or the true proportion of women faculty) in a way that corresponds to adding the recorded "favorable" and "unfavorable" observations to the imagined pairs of observations encoded in her prior distribution. For example, if someone's prior distribution corresponds to 2 observations of heads and 3 observations of tails, and she then sees 3 actual heads tosses and 17 actual tails tosses, then her *posterior distribution* will correspond to 2+3=5 "heads" tosses and 3+17=20 "tails" tosses (the scare quotes here indicate that we are summing the imagined tosses and real tosses). More generally, if someone's prior distribution corresponds to *a* favorable and *b* unfavorable observations, and she then makes *c* favorable and *d* unfavorable further observations, then her posterior distribution will correspond to *a*+*c* favorable and *b*+*d* unfavorable observations.

The distribution of someone's credences over hypotheses about the true value of an unknown quantity determines what she should expect the next observation to be. In our examples, your credences in the myriad hypotheses about the weighting of the coin determine how likely you should think it is that the first toss will land heads, and your credences in the hypotheses about the true proportion of faculty among women employed at universities determine how likely you should think it is that the next woman employed at a university who you meet will be faculty. To return to our toy example, if you think it is 20% likely that 60% of women employed at universities are faculty, 50% likely that 40% are faculty, and 30% likely that 10% are faculty, then your credence that the next woman employed at a university who you meet will be faculty should be (0.2*0.6)+(0.5*0.4)+(0.3*0.1)=0.35, so you should expect that she will not be faculty — that is to say, you should have a low credence in her being faculty. With a continuous distribution of credences over a set of hypotheses about the true value of the unknown quantity, like those depicted in Fig. 1, the general idea is the same, but the associated summation is slightly different as there are infinitely many "hypotheses" about the true underlying proportion, so we integrate over them.

When Boris makes his estimate about Zoë's coin toss, his estimate of the probability that it will land heads should be a *weighted average* of the possible biases of the coin to which he assigns some credence, with each one "weighted" (i.e. multiplied) by his credence that this is the coin's true bias. Likewise, in **Gender Bias Study** your estimate of the probability that the next woman employed at a university who you meet will be a faculty member should be a weighted average of the possible proportions of faculty among women university employees to which you assign some credence, with each one weighted by your credence that this is the true proportion. These estimates are known as the agent's *expectations* of the true value of the unknown quantity. The agent's expectation is the mean of their probability distribution. This quantity exists before the agent observes any evidence (the prior mean), and a different one exists after the agent updates on some new evidence (the posterior mean). In each case, it is a weighted average, but the distribution with respect to which it is evaluated changes in response to the evidence.

[4] Of course it would also correspond to real observations, but when there are no real observations we use the fiction of imagined observations to characterize the prior distribution.

As we noted above, the natural-seeming answer is wrong to think that, without any other pertinent information, epistemic rationality requires that your estimate of the overall proportion of women employed at universities who are faculty must be the proportion reported in the study. Reading about the study is like observing some tosses of a coin; rationality requires that you incorporate this new information into the distribution of your credence over the myriad hypotheses about the true proportion of faculty members among women employed at universities. But which distribution results from this process depends on what your prior distribution over these hypotheses was. Agents with different prior distributions will end up with different posterior distributions after reading about the study.

To start with a uniform distribution, such as that in the left-panel of Fig. 1, is equivalent to having prior credences that correspond to having "observed" precisely one imaginary heads toss and one imaginary tails toss. Uniformity is as "flat" as our credal distribution can get, so to speak. This is close to the natural-seeming answer, but not the same, because even a uniform prior will exert a slight influence on the posterior. So we can ask: is there any prior that will exert zero influence on the posterior? This is the prior required by the natural-seeming answer. Mathematically, this can be constructed: what we need is a prior corresponding to the very specific information state of zero "favorable" observations and zero "unfavorable" observations. But epistemic rationality does not require us to have such a prior in the absence of any information about some topic. Indeed, epistemic rationality *forbids* such a prior, as it violates the probability axioms. This is because such a prior will assign infinite weight to very small and very large values of the true unknown bias of the coin or proportion of faculty, with the result that the sum of our credences will not be 1. But the defect is greater than that. Suppose we have a credence-function which is such that the credences in all the possibilities add up to, say, 3. Then we can take this credence-function and divide it by 1/3 in order to arrive at a probabilistically coherent credence-function. In this case, 1/3 is called the *normalizing constant*. But when the sum of a credence-function is infinite there is no normalizing constant that would turn it into a valid credence-function. The epistemic state required by the natural-seeming answer is one according to which the agent's credence in very small and very large values of the unknown quantity is infinitely greater than her credence in middling values. Such an agent would happily accept very absurd bets about the coin's bias or the true proportion of faculty.

So, epistemic rationality clearly does not require that your estimate of the overall proportion of women employed at universities who are faculty must be the proportion reported in the study. On the contrary, such an estimate under these circumstances corresponds to a prior distribution that violates the basic axioms of probability. Far from being required, this prior distribution is prohibited by the most elementary requirements of epistemic rationality. So the Frequency-Credence Connection cannot be a genuine principle of epistemic rationality.

We could try to salvage the natural-seeming answer by adopting its closest relative — the idea that one must have a prior credal distribution corresponding to one imagined "favorable" observation and one imagined "unfavorable" observation. As we briefly mentioned above, this is a uniform distribution, depicted in the left-panel of Fig 1a. This corresponds to a famous principle in probability theory known as the "Laplacean Rule of Succession" (see Laplace 1814, Zabell 2005, Huttegger 2017).

But such a credal distribution just represents another very specific information state: exactly one favorable and one unfavorable observation. In the absence of any pertinent information on the true proportion of faculty among women employed at universities, this is no more warranted by epistemic rationality than any other prior corresponding to any other maximally specific information state.

In the next section we discuss a way to settle the tie between priors corresponding to very specific information states: we hold that the tie is broken by the agent's attitude toward *epistemic risk*.

b. *A positive proposal: Epistemic Risk as a Guide*

Recall (from §2.2) that a good scoring rule must be truth-directed, continuous, and strictly proper. This does not narrow things down very much, as there are infinitely many functions that satisfy the accuracy checklist. A common example of a scoring rule is the "Brier score" – named after Glenn Brier, a meteorologist who proposed the metric as a way of measuring the accuracy of weather forecasts – which takes inaccuracy to be the squared Euclidean distance between an agent's credence and the truth-value of the proposition . In other words, if it rains the inaccuracy of the agent's credence for rain *x* is given by (1- *x*)^2, whereas if it does not rain it is given by (0 - *x*)^2. This in keeping with the spirit of least squares estimation. However, no tenet of epistemic rationality requires that we weigh mistakes in the false positive direction (i.e., high confidence in rain when it does not rain) equally with mistakes in the false negative direction (i.e., low probability for rain when it rains). Indeed, a scoring rule of the form *y*(1- *x*)^2 if it rains and *z*(0- *x*)^2 if it does not would still be truth-directed, continuous, and strictly proper for any positive constants *y* and *z*.

These different scoring rules embody different *attitudes toward epistemic risk*. An agent's attitude toward epistemic risk depends on her assessment of the relative disvalue of the two types of epistemic error: high credence in a falsehood and low credence in a truth. Different credences seem more or less "risky" to her, depending on how she evaluates these types of error. For example, a weather forecaster might be more worried about having a low credence in a truth than having a high credence in a falsehood if the proposition in question is that there is a tornado nearby — a false alarm might be inconvenient, but failing to predict a tornado would be disastrous. So, in this case, it is quite reasonable for the weather forecaster to see false negatives as worse than false positives. Moreover, as this example shows, assessments of the disvalue of the two types of epistemic error need not be based solely on the agent's intrinsic aversion to having high credences in truths and low credences in falsehoods. The forecaster does not think that false negatives are inherently worse than false positives. Rather, her asymmetric attitudes to the costs of error are based on her pragmatic concerns. But this is fully compatible with the requirements of epistemic rationality.

We can express this by saying that a Bayesian agent possesses an *epistemic risk function*, which constrains how she evaluates inaccuracy. For every credence between 0 and 1, her epistemic risk function determines how risky that credence is. The least risky credence is the one where she is guaranteed a certain inaccuracy score no matter whether the proposition in question turns out to be true or false. To illustrate this idea, return to the **Gender Bias Study**. Imagine two agents, A and B, each under the circumstances described. Before reading about the study, A is indifferent between false positive and false negative mistakes about Mary; she wants to be accurate, and has no preference between falsely assuming that Mary is an administrative assistant and falsely assuming that she is a faculty member. B, meanwhile, thinks that falsely assuming that Mary is an administrative assistant is much worse than falsely assuming that she is a faculty member in a world in which faculty members are held in higher regard. Here A and B differ in their attitudes toward epistemic risk with respect to propositions regarding Mary's role. B, like the weather forecaster, thinks that one type of error is worse than the other in his circumstances, and thus that low credence is riskier than high credence for the proposition "Mary is a faculty member" whereas high credence is riskier than low credence for the proposition "Mary is an administrative assistant".

Suppose that A and B's epistemic risk functions are given by the following curves, where the *x*-axis represents a credence, and the *y*-axis represents their assessment of the riskiness of the credence:

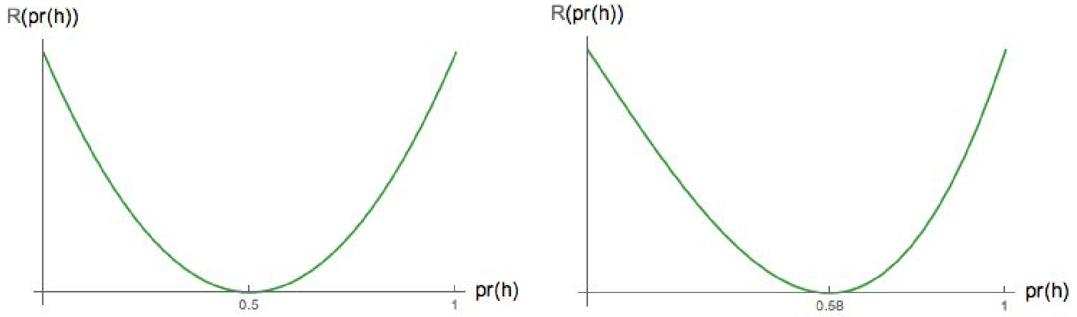

Fig 2. *Symmetric epistemic risk function (left-panel), and non-symmetric epistemic risk function (right panel)*

These curves are consistent with the description provided so far of A and B's attitudes to epistemic risk.

In previous work, one of us has shown that if an epistemic risk-function is convex and continuous – as these are – then, drawing on a result from Savage (1971), we can derive from it a unique scoring-rule that meets the accuracy checklist (Babic 2018). Here A's risk-function corresponds to the Brier score, whereas B's risk-function corresponds to a different scoring rule: another quadratic score, but one that increases differentially in the false positive and false negative error directions, with added "penalty points" for increases in inaccuracy in the false positive direction. The associated scoring rules are depicted in Figure 3, with the agent's credence on the x-axis and the inaccuracy score of this credence on the y-axis. (What score a credence gets depends on whether the proposition in question is true or false, so we have included both a solid blue line representing the score if the proposition is true and a dashed red line representing the score if it is false.) For reference, we have also plotted the agent's risk functions from Figure 2.

Thus, there is a close relationship between the agent's assessment of epistemic risk and the associated scoring rule. The shaded pink and grey areas in Figure 3 display the verdicts of the agent's epistemic risk function. We have included the riskiness of a credence of 0.2 in a proposition and one of 0.8 in the same proposition to illustrate how someone's assessment of the riskiness of a credence varies as she assesses the severity of different errors differently. In the left-panel a credence of 0.2 is equally as risky as a credence of 0.8, whereas in the right-panel a credence of 0.2 is barely risky at all, whereas a credence of 0.8 is extremely so. This is because the agent depicted in the right-panel is especially concerned about making mistakes in the false positive direction, whereas the agent depicted in the left-panel is indifferent between the two types of error.

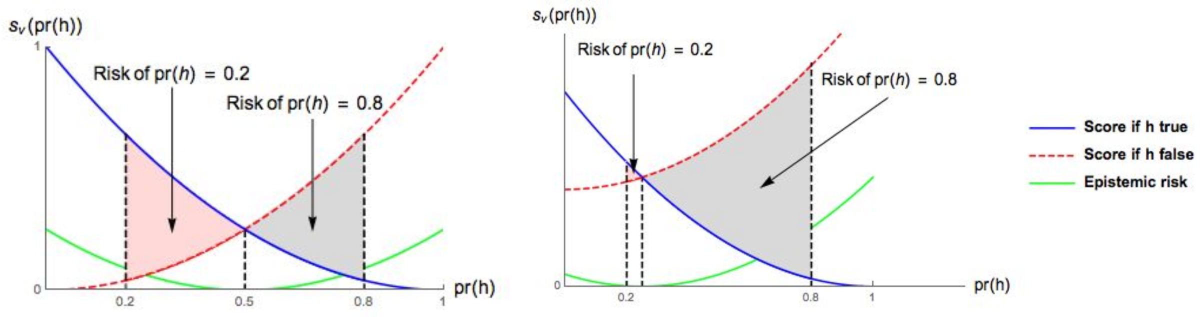

Fig. 3. *Symmetric scoring rule and epistemic risk (left panel), non-symmetric scoring rule and epistemic risk (right panel). The green (risk) curve encodes the shaded areas (pink and gray).*

A's scoring rule is *symmetric*, since she holds that false positive mistakes are equally as bad as false negative mistakes (i.e., since her associated epistemic risk function is symmetric), whereas B's is *non-symmetric* (since her associated epistemic risk function is non-symmetric). But, to repeat, the choice between particular scoring-rules such as these is not settled by the requirements of epistemic rationality. Both of these scoring rules satisfy the checklist of epistemic desiderata: they are continuous, monotonic, and strictly proper.

Since epistemic rationality is silent on which particular scoring rule is required, this choice may be rationalized by the agent's attitude toward epistemic risk. Each epistemically permissible scoring-rule embodies a different attitude toward the risk of graded error, and may be justified on this basis.

Suppose that A and B read about the **Gender Bias Study**. They learn that a seemingly well-designed study, reported in the Washington Post, found that 70% of women employed at academic institutions had administrative roles, while 30% were faculty members. After learning this, A and B's credences will change, since the information reported in the study is clearly relevant to their estimation of the proportion of women employed at universities who are faculty. We assume that they update by Bayesian conditionalization.

What should A and B think about the proposition "Mary is an administrative assistant"? In general, they should adopt whichever credences minimize their expected inaccuracy. Before they read about the study, they have (by hypothesis) no information about the probability that Mary is an administrative assistant, so there is nothing in the epistemic toolkit to help them determine what credence they should hold. However, they can use their attitude to epistemic risk to identify an appropriate prior.

Consider A, first. If we assume that she seeks to minimize epistemic risk, her prior credence, before observing any evidence, that Mary is an admin should be 0.5. These attitudes to epistemic risk require a prior probability for the true proportion of faculty that is uniform. (Recall that a uniform distribution is equivalent to a distribution with one imaginary observed faculty member and one imaginary observed administrative assistant.)

This distribution is forced by the agent's measure of epistemic risk and her desire to minimize it in adopting a prior. This is because the uniform distribution for the unknown proportion is the only distribution which is such that for any possible true value of the proportion (i.e, for any possible hypothesis), an error in the direction of that hypothesis is treated the same as an error in the direction of any other hypothesis.[5]

After reading about **Gender Bias Study**, A will update by Bayesian conditionalization. And she will formulate a credence about Mary — the predictive inference — which will correspond to the mean of her posterior distribution. Recall that, in the sorts of cases that we are considering, an agent's prior can be accurately described with a pair of numbers (a, b) corresponding to "imaginary" successful/failed observations, the data can be described with a pair of numbers (c, d) corresponding to observed successes/failures, and the agent's posterior can thus be described with the pair of numbers (a+c, b+d). To refer to the unknown proportion of faculty or unknown bias of the coin, we will use the symbol $\boldsymbol{\theta}$. In other words, the distributions given by the pairs of numbers, as just described, are distribution about $\boldsymbol{\theta}$. And we can conveniently compute Bayesian updating by combining imaginary observations with real observations. Thus, A's posterior credence for $\boldsymbol{\theta}$ after becoming aware of the Gender Bias Study must be (1 + 150, 1 + 350).

[5] For example, consider the approximately bell-shaped curve from the right-panel of Fig 1. If an agent were to adopt this curve as her prior, it would suggest that she is extra sensitive about errors for values near the peak of that curve. The only curve where she is equally sensitive to error with respect to every hypothesis is the flat or uniform curve -- that is, the curve depicted in the left-panel of Fig 1.

We will introduce one more mathematical fact that will be a helpful heuristic for us. We have said that the prediction that Mary is an admin is based on the mean of the posterior distribution. If we have a distribution for $\boldsymbol{\theta}$ given by a pair of numbers (a, b), where $\boldsymbol{\theta}$ can be any real number between 0 and 1, inclusively, then its mean, under very general conditions that are satisfied here, will be a/(a+b). Thus, A's credence for the proposition that Mary is an admin is given by 351/(151 + 351) = 0.699. Here A is quite confident that Mary is an administrator, but just slightly less so than she would have been had she applied the naive Frequency-Credence Connection principle — which would suggest the sample mean as the credence, i.e., 350/500 = 0.7.

Now consider B. Her safest prior credence function, before seeing any evidence, in the proposition that Mary is an administrator is 0.25. This is where her epistemic risk function, as depicted in the right panel of Figure 1, reaches its minimum. This epistemic risk function requires a prior for $\boldsymbol{\theta}$, given by (a, b) which is such that a/(a+b) = 0.25. This is due to the mathematical fact introduced above. Unlike A, whose epistemic risk attitudes required the specific (1, 1) uniform prior, B's attitudes to epistemic risk are more permissive, so to speak. They are compatible with any prior (a, b) provided that a/(a+b) = 0.25. We have not said anything else about B's attitudes to epistemic risk that would further constrain her set of permissible priors for $\boldsymbol{\theta}$. Since the posterior is given by (a + 350, b + 150), and we know that a/(a+b) must be equal to 0.25, suppose that for B, a = 100 and b = 300. This satisfies everything we have said about B's attitudes to epistemic risk. And it leads to a posterior distribution corresponding to (450, 450), again, each number corresponding to the sum of imagined and real admins and imagined and real faculty. B now has to formulate a credence as to whether Mary is an administrator. While she is not going to use Laplace's Rule of Succession, since her prior for theta is not uniform, she will still use the mean of her posterior distribution. Huttegger (2017) refers to this principle for making predictive inferences as the Generalized Rule of Succession -- i.e., it is the principle which requires the agent to use the mean of her posterior distribution in order to make predictions, but where her prior need not be uniform. Using this principle, her credence that Mary is an administrator is 450/900 = 0.5.

In Figure 3, below, we depict the situation described in the previous two paragraphs. A's prior for $\boldsymbol{\theta}$ is given by the green figure in the left panel, whereas her posterior is given by the red curve. That is, she moves, as the arrow indicates, from the green curve to the red curve after updating her credences for $\boldsymbol{\theta}$ on the **Gender Bias Study**. Her credence regarding the proposition that Mary is a faculty member is given by the mean of the green curve, before seeing the study (0.5), and by the mean of the red curve, after seeing the study (0.31). Meanwhile, B's credences for $\boldsymbol{\theta}$ are given by the green and red curves in the right panel (before and after seeing the study, respectively) and B's credence for the proposition that Mary is a faculty member is likewise given by their mean (0.75 before seeing the study and 0.5 after seeing the study).

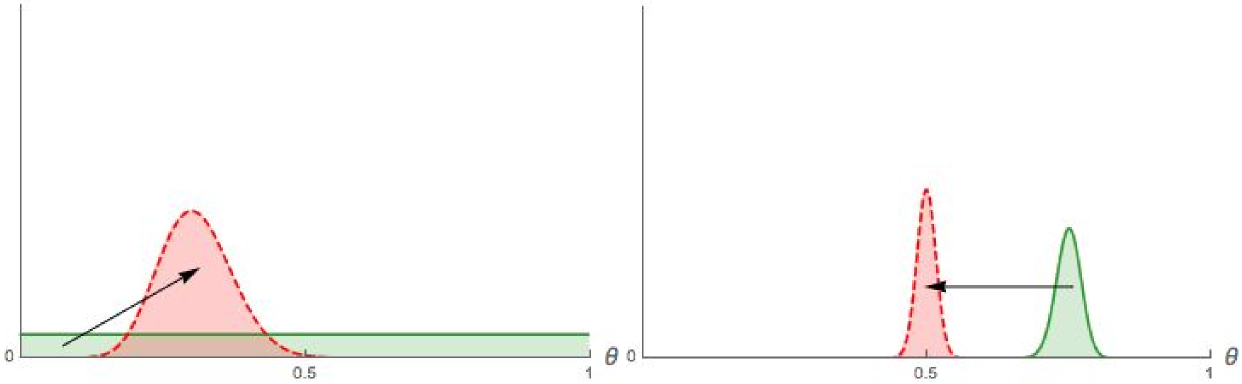

Fig 4. *The left-panel depicts the prior (green curve) for A and the posterior (red curve) for A. The right-panel depicts the prior (green curve) for B and the posterior (red curve) for B. The arrows indicate Bayesian conditionalization on the Gender Bias Study.*

As the diagrams suggest, A ends up quite confident that Mary is not a faculty member (moving from 0.5 to 0.31 credence in that proposition) while B ends up rather indifferent (moving from 0.75 to 0.5 credence in that proposition). Thus B avoids assuming that Mary is an administrative assistant based on the statistical information presented in the gender bias study. And this is so even though B updates by conditionalization, keeps his credence-function coherent, and uses a truth-directed, continuous, strictly proper scoring rule to measure inaccuracy.

We would like to anticipate a common objection here. It may seem odd that B started more confident in the proposition that Mary is a faculty member than in the proposition that she is not. Should she not be indifferent in the absence of information? No. This is precisely what we deny. The notion that an agent must be indifferent because they are ignorant is self-defeating. A 0.5 credence in the proposition that Mary is a faculty member presupposes a uniform credence for $\boldsymbol{\theta}$. This in no way adequately represents an absence of information. Instead, it captures a highly specific information state — one in which the agent deems every value of theta to be exactly as probable as any other. But in the absence of information, this distribution regarding $\boldsymbol{\theta}$ is by hypothesis no more justified than any other distribution. In the absence of information, epistemic rationality is silent, which is why we turn to the agent's normative attitudes to epistemic risk for assistance in formulating a prior that adequately captures their attitudes to error in the situation they find themselves in. B's epistemic behavior reflects his attitudes toward the moral and political costs of error in this type of case. And he has struck this balance differently than A has, deeming that false confidence in the proposition that Mary is an admin is worse than false confidence in the proposition that she is faculty, given the way these occupations are socially regarded in our society. From an epistemic point of view, both ways of balancing the error costs are appropriate.

c. *Avoiding the Sad Conclusion*

We have seen that it is the agent's attitude toward epistemic risk that rationalizes her policy for predictive inference. Epistemic rationality requires agents to make such inferences so as to minimize their expected inaccuracy, as measured by a permitted scoring-rule – that is to say, a truth-directed, continuous, strictly proper scoring-rule. But the requirements of epistemic rationality do not pin down a privileged scoring-rule. This is up to the agent to determine, according to her attitude toward epistemic risk. The agent's attitude toward epistemic risk fixes the shape of her measure of inaccuracy, and thus determines how she should make predictive inferences. Moreover, the agent's attitude toward epistemic risk embodies a normative assessment: an assessment of the badness of having a high credence in a falsehood or a low credence in a truth. And this assessment need not be made on purely alethic grounds. Just as a weather forecaster might be more concerned with false negative mistakes when the proposition in question is "There is a tornado nearby", so too may B be more concerned with false positive mistakes when the proposition in question is "Mary is an administrative assistant". The forecaster knows that failing to prevent multiple deaths is much worse than mildly inconveniencing people, and her attitude toward epistemic risk is sensitive to the greater moral severity of this sort of mistake. Likewise, B believes that he would wrong Mary by falsely assuming that she is an administrative assistant, but not by falsely assuming that she is a faculty member, and his attitude toward epistemic risk is sensitive to the greater moral severity of this sort of mistake. In the absence of information, these attitudes to epistemic risk recommend an appropriate prior, which in turn affects where the two agents end up after reading about the study.

We see no reason why pragmatic, moral, and political concerns should not be used to determine the relative badness of the two types of error. Indeed, *some* assessment of non-alethic badness must presumably be used to settle the choice of a specific scoring rule, since alethic considerations drastically underdetermine this choice. The familiar refrain that we should value high credence in truth and low credence in falsehood, and should disvalue high credence in falsehood and low credence in truth, says nothing as to the *relative* value of these things. But any particular measure of inaccuracy takes a view on the relative value of these things. Even the traditional Brier score reflects the implicit assumption that the costs of false negative and false positive mistakes are equally severe. No scoring-rule is normatively neutral.

This helps us to avoid the Sad Conclusion. To repeat, this is the Sad Conclusion (*op. cit.*):

> As long as there's a differential crime rate between racial groups, a perfectly rational decision maker will manifest different behaviors, explicit and implicit, towards members of different races. This is a profound cost: *living in a society structured by race appears to make it impossible to be both rational and equitable*.

The Sad Conclusion is false. Agent B is a counterexample: he is both rational and equitable. He has done nothing to violate the requirements of epistemic rationality, yet he exhibits precisely the sort of reluctance to draw a pernicious predictive inference about Mary that is intuitively morally praiseworthy. Meanwhile, someone who stands ready to draw pernicious predictive inferences based on sample data is not a paragon of tough, clear-headed rationality. They might be epistemically irrational, if one of the epistemic criticisms of pernicious predictive inference from 2.1 holds true of them. But even if none of these criticisms hold, the attitude toward epistemic risk that this person embodies is not required by epistemic rationality. On the contrary, it is just one of many epistemically permissible attitudes toward epistemic risk. Such an agent therefore has to do the hard normative work of explaining why their attitude to epistemic risk is the right one to adopt. Here the deck is stacked against them, since this attitude seems morally and politically objectionable in our current social context. It is objectionable to hold that false positive and false negative mistakes are equally disvaluable in the sort of case under discussion, since they are *not*, in fact, equally disvaluable — to adopt this attitude toward epistemic risk is to be indifferent toward all of the moral and political costs of error.

In sum: our view is that the problem with drawing pernicious predictive inferences, and with valuing false positive and false negative errors equally in the associated cases, is not that doing so makes you epistemically irrational, but just that it makes you a jerk.

4. Accepting statistics

The above concerns predictive inferences: inferences about the probability that an individual has a certain trait, based on information about the prevalence of the trait within a sample of a population to which she belongs. But drawing such predictive inferences is not the only thing that someone might be epistemically required to do upon encountering a statistic. One alternative is that she may be epistemically required to *accept the statistic*. For instance, in **Gender Bias Study**, she may be epistemically required to accept that of people employed at academic institutions, 70% of women are administrative assistants and 30% are faculty, whereas for men, things are the other way around.

We have emphasized that there is no epistemic requirement to assume that the overall frequency of a trait in a population is equal to the frequency that one hears that a study found in a sample of that population. So there is no epistemic requirement to accept statistics in this sense. But we think that it is plausible that some related epistemic requirements hold. The agent in **Gender Bias Study** cannot simply ignore the information that she hears, revising none of her doxastic states. On the contrary, she has gained some new evidence, and she must decide what her evidence is and then respond accordingly. If there is reason to doubt the study's validity — perhaps because it used a small sample, or because it failed to investigate covariants — then she may take her evidence to be that *a study* *found evidence of* a certain population frequency, rather than that *there is* a certain population frequency. If there is reason to doubt the newspaper's reliability — perhaps because it has a track record of inaccurately reporting scientific studies — then she may take her evidence to be that *a newspaper reported that a study found evidence of* a certain population frequency. And if she has reason to be even more skeptical, then she might take her evidence to be that *she had a series of sensory experiences as of a newspaper reporting that a study found evidence of* a certain population frequency. But she has to start somewhere. This is simply how responding to evidence works, on the accuracy-based approach to epistemology; this approach requires that we update our credences by conditionalizing on what we have learned, and conditionalization begins with the agent identifying something that she has learned.

One might take this picture of epistemic rationality to support the Sad Conclusion after all. But that would be a mistake. This picture supports the Sad Conclusion only if there are statistics that someone is epistemically required to accept, but also morally prohibited from accepting. And we doubt that there are any of these. Indeed, when statistical facts about morally or politically problematic phenomena are reported from credible sources, accepting them may be morally *required*. It is an unfortunate fact that there are enormous racial disparities in incarceration rates in the contemporary United States, and this fact reflects underlying structural injustices to which one should not be blind. Similarly, when there is reason to distrust the source of a piece of statistical information, someone may still be morally required to accept that the source reported the information. We are morally required to resist the structural social injustices that underlie and explain these statistical facts, and we cannot do so effectively without informing ourselves of the facts — both statistical facts about the prevalence of pernicious traits in social groups, and further facts about how these topics are publicly reported. For example, it is unlikely that we will be able to effectively address the structural problems underlying incarceration rates among African-Americans without accepting statistical information about what those rates are. Similarly, it is unlikely that we will be able to address these structural problems if we ignore information about how the rates are reported, particularly if public reports are misleading in some way. So, the Sad Conclusion is false again: no tension between the requirements of epistemic rationality and those of morality lurks in these waters. On the contrary, the sort of epistemic behavior that might be rationally required — accepting information about the prevalence of pernicious traits within social groups, and/or information about how this information is reported — is at least morally permitted, and perhaps morally required.

5. Direct inference

We have so far been discussing predictive inferences: inferences about the probability of an individual's possessing a trait based on data about the prevalence of this trait within samples taken from populations — in our cases, social groups — to which she belongs. In such cases, the individual herself is not part of the sample that was formerly used to determine the data. When the individual herself was part of the sample, the kind of inference performed is often called *direct inference*. Here is a variant on our case that makes it a case of direct inference:

> **Gender Bias Study 2**. One morning you read a report in the Washington Post about a study reporting gender discrepancies in academic employment in the United States. The authors of the study surveyed 1,000 people, 500 men and 500 women. They found that 70% of women held administrative roles and 30% held faculty roles, whereas the opposite was true of men. The study seems to have been conducted competently; the authors stipulated in advance that the survey would end after 500 men and 500 women responded, they did not perform multiple comparisons of their data in order to find evidence of gender bias, they disclosed all covariates that were tracked, and no observations were excluded. Before reading this article, you had no information about the relative distribution of women and men across different occupational roles in academia. After reading about the study you meet Mary, a new neighbor who tells you that she is employed

> at your local university and that she participated in the study that was just reported in the Washington Post. Mary does not tell you in what capacity she is employed.

Direct inference is importantly different from predictive inference. It is a kind of inference for which the frequency-credence connection theoretically does hold, and, most importantly for present purposes, a kind of case in which a version of the Sad Conclusion theoretically could arise.

The frequency-credence connection holds for direct inference under some very specific circumstances, originally discussed by Levi (1977). Suppose that you learn some data about the prevalence of a trait in a population, and you have no further evidence whatsoever that makes it any more or less likely that any particular member of this population possesses the trait than any other member. If you are in this situation, then your credences about the members of the population are said to be *exchangeable*. Under conditions of exchangeability, if you are asked to estimate the probability that a randomly-selected member of the population possesses the trait, then a version of the frequency-credence connection holds: your estimate of the probability that the randomly-selected individual possesses the trait should be equal to its frequency in the population. In this case, being slow to assume anything about the probability that the randomly-selected individual possesses a pernicious trait would violate even the formal coherence requirements of epistemic rationality. So, if being slow to assume anything pernicious about the randomly-selected individual is morally or politically required even in this kind of case, then a version of the Sad Conclusion does hold.

However, we are not convinced that being slow to assume that individuals possess pernicious traits in direct inference under conditions of exchangeability really is morally required. To see why, note firstly that we are almost never under conditions of exchangeability, and moreover that we are not under such conditions in any of the kinds of case that moral philosophers, epistemologists, and legal scholars have discussed in their respective literatures. The cases of interest are those involving real-life encounters with particular individuals who one knows is a member of a population within which a trait was found to have a certain frequency. But real-life encounters do not occur under conditions of exchangeability. As soon as you meet someone, you gain a tremendous amount of information about them — about their appearance in the first instance, and then about their mannerisms and patterns of speech, their accent, the content of what they say to you, and so on, as you go on to interact. These pieces of information can all be relevant to the probability that they possess the trait in question. So, once you learn them, it is no longer true that you have no further evidence about any member of the population that makes it more or less likely that they possess the trait in question than any other member. On the contrary, you have a great deal of evidence about the person you just met that is pertinent to the question of whether they possess the trait, and that thereby sets them apart from other members of the population. In other words, once you actually meet somebody from the population about whom you know a statistic regarding the prevalence of a trait, you are no longer under conditions of exchangeability. But the frequency-credence connection only holds for direct inference under conditions of exchangeability. So it does not hold for real-life encounters of the sort that trouble moral philosophers, epistemologists, and legal scholars.

The kind of case for which the frequency-credence connection holds, then, is an unusual sort of case. It is a case in which we are asked to estimate the probability that a randomly-selected person possesses a trait, given that they belong to a population in which it has been determined that a certain percentage of people possess the trait, and given *no* other information about the individual besides this fact (Levi 1977). It is not at all clear that we are morally required to resist conforming our credence to the population frequency in such a case. In so doing, we cannot fairly be accused of failing to respond to the particular individual as an individual, since we have been deprived of any information about her individuality — this is exactly what it is to be under conditions of exchangeability. Moreover, when adopting credences in propositions about randomly-selected people in this way, we must do so in the abstract, without meeting the people and in ignorance of their identity — since, if we met them or knew anything about their identity, then we would not be under conditions of exchangeability. This is Levi's random selection constraint, which is misunderstood in White (2010). It is extremely unclear that there is any sense in which we wrong a person by drawing the direct inference about them, considered in the abstract, in such a case. If there is a sense in which we wrong them, it remains to be seen what it is.

REFERENCES

Babic, Boris (2018). "A Theory of Epistemic Risk." *Philosophy of Science* Forthcoming.

Basu, Rima. and Schroeder, Mark (2018). "Can Beliefs Wrong?" In *Philosophical Topics* (Special Issue). Forthcoming.

Basu, Rima (*ms*). "The Moral Stakes of Racist Beliefs".

Buchak, Lara (2014). "Belief, Credence, and Norms." Philosophical Studies 169(2):285-311.

Easwaran, Kenny (2013). "Expected Accuracy Supports Conditionalization -- and Conglomeration and Reflection." *Philosophy of Science* 80(1): 119-142.

Gardiner, Georgi (*ms*). "Evidentialism and Moral Encroachment".

Gendler, Tamar Szabo (2011). "On the Epistemic Costs of Implicit Bias." *Philosophical Studies* 156 (1), 33-63.

Huttegger, Simon (2017). *The Probabilistic Foundations of Rational Learning*. Cambridge University Press.

Kyburg, Henry (1974). The Logical Foundation of Statistical Inference. Dordrecht: Reidel.

Leitgeb, Hannes and Richard Pettigrew (2010). "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77(2): 236-272.

Levi, Isaac (1977). "Direct Inference." *The Journal of Philosophy*. 74(1): 5-29.

Munton, Jessie (*ms*). "The Epistemic Flaw with Accurate Statistical Generalizations"

Moss, Sarah (2016). *Probabilistic Knowledge*. Oxford, UK: Oxford University Press.

Reichenbach, Harry (1938). *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago University Press.

Thomson, Judith Jarvis (1986). "Liability and Individualized Evidence." Law and Contemporary Problems 49(3):199-219.

Tribe, Lawrence (1971). "Trial by Mathematics: Precision and Ritual in the Legal Process." Harvard Law Review 84(6): 1329-1393.

White, Roger (2010). "Evidential Symmetry and Mushy Credence." In T.S. Gendler and J. Hawthorne (eds), *Oxford Studies in Epistemology*, Vol 3, 161-186. Oxford University Press.

Zabell, Sandy (2005). *Symmetry and its Discontents.* Cambridge University Press.