A Theory of Epistemic Risk

Boris Babic*†

I propose a general alethic theory of epistemic risk according to which the riskiness of an agent's credence function encodes her relative sensitivity to different types of graded error. After motivating and mathematically developing this approach, I show that the epistemic risk function is a scaled reflection of expected inaccuracy (a quantity also known as generalized information entropy). This duality between risk and information enables us to explore the relationship between attitudes to epistemic risk, the choice of scoring rules in epistemic utility theory, and the selection of priors in Bayesian epistemology more generally (including the Laplacean principle of indifference).

1. Introduction. My goal in this article is to develop and defend a general theory of epistemic risk within the epistemic utility framework. In light of the growing influence of decision-theoretic approaches to epistemology, it is natural to consider what role risk will play in the normative assessment of an agent's credence function. To date, there is very little literature on this topic.[1]

Received January 2018; revised May 2018.

*To contact the author, please write to: California Institute of Technology, MC 101-40, 1200 E. California Blvd., Pasadena, CA 91125; e-mail: bbabic@caltech.edu.

†I would like to thank Jim Joyce for his invaluable support and guidance in developing this project. I have also received helpful feedback from Sara Aronowitz, Gordon Belot, Daniel Drucker, Frederick Eberhardt, Dmitri Gallow, Rich Gonzalez, Hilary Greaves, Christopher Hitchcock, Simon Huttegger, Sarah Moss, Matt Parker, Richard Pettigrew, Peter Railton, Julia Staffel, Anubav Vasudevan, Brian Weatherson, and audiences at the University of Michigan, University of Chicago, London School of Economics, and California Institute of Technology. Special thanks are also due to two anonymous reviewers from *Philosophy of Science* for their very helpful comments and feedback. Research for this project was supported by the Social Sciences and Humanities Research Council of Canada.

1. Related articles include Levi (1962) (exploring the cost of mistakes for categorical belief), Fallis (2007) (focusing on the value of experiments), Buchak (2010) (evaluating the effect of a nonexpected utility measure of risk aversion on evidence gathering), Pettigrew (2016a) (employing a minimax decision rule to defend the principle of indifference), and Pritchard (2017) (defending a modal account of risk for traditional approaches in epistemology). However, none of the cited papers offer (or aim to offer) a general theory of epistemic risk for degrees of belief.

Philosophy of Science, 86 (July 2019) pp. 522–550. 0031-8248/2019/8603-0007$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

I make some fairly basic assumptions about epistemic value. For illustrative purposes, we will proceed under the fiction that an agent's selection of a credence function may be treated as an epistemic act such that the rationality of that act can be evaluated using the tools of ordinary decision theory. One epistemic act is preferable to another if it increases expected epistemic utility, where epistemic utility is given in terms of a scoring rule. Scoring rules measure the accuracy of an agent's credence function. Thus, accuracy is our primary epistemic commodity.

If we are to develop a theory of risk within the framework of epistemic utility, the riskiness of a credence function should reflect an agent's exposure to potential losses in accuracy. The theory should be alethic, so to speak. We will see that following this line of inquiry, the shape of an agent's epistemic risk function reflects her relative sensitivity to different types of graded error. In simple cases, this implies that the shape of the risk function reflects an agent's attitude toward the relative cost of increasing inaccuracy in the direction of false positive (type I) mistakes against the cost of increasing inaccuracy in the direction of false negative (type II) mistakes. On larger sample spaces, the risk function reflects an agent's attitude to increasing inaccuracy in the direction of every possible outcome. Meanwhile, the curvature of the risk function encodes attitudes toward marginal changes in accuracy and local sensitivity to error.

To develop a measure of epistemic risk that captures the preceding idea, I propose an approach inspired by Rothschild and Stiglitz's (1970) measure of economic risk in terms of stochastic dominance. On my approach, one credence function is riskier than another if it is a mean-preserving spread of it, and the least risky credence function is the one that guarantees a particular accuracy score regardless of the state of the world. We will see that for credence functions, mean-preserving spreads in accuracy are equivalent to certain changes in expectation and that a plausible measure of risk, therefore, is the difference in expectation from the risk-free credences. In simple cases, this measure has a very natural interpretation in terms of the difference between the agent's best and worst outcomes.

Following Grunwald and Dawid (2004), I use the term 'general entropy' to refer to the expected accuracy of a credence function evaluated with respect to itself (we will see why below). As a result, the notion of epistemic risk I advocate is also a measure of entropic change. Indeed, the main formal contribution of this article is a duality theorem connecting epistemic risk and general entropy, which will be established in section 5: namely, that under very general conditions risk is a scaled reflection of general entropy. That is,

$$\text{risk} + \text{entropy} = k.$$

This is a fruitful link between risk and information entropy. From every risk function we may derive a unique scoring rule, and the agent's attitude to dif-

https://doi.org/10.1086/703552 Published online by Cambridge University Press

ferent types of error will determine the shape of her score. For example, if she considers the different error costs to be equal, her score will evaluate equally changes in accuracy in the direction of each outcome. If such an agent seeks to minimize epistemic risk, she will identify a uniform prior by applying the Laplacean principle of indifference (POI). However, the uniform prior minimizes epistemic risk only if the different types of error are treated equally. This is quite a substantial assumption, and a version of it appears in Pettigrew's (2016a) accuracy argument for the POI. More generally, the relationship between risk and general entropy suggests that there exists a family of indifference principles (rather than a unique POI) each reflecting a different way of evaluating the error costs of a prospective credence function. This highlights the normative commitments that come with endorsing an uninformative or flat prior. The agent's risk profile, therefore, is in an important sense epistemically central. Once we know what it is, we can determine the appropriate measure of risk, the associated entropy, and the scoring rule.

The article proceeds as follows. Section 2 provides an overview of competing approaches one might use to develop a theory of epistemic risk, focusing in particular on the difference between alethic and modal conceptions of risk. In section 3, I describe the relevant formal concepts. In section 4, I develop the theory of epistemic risk for a simple case. In section 5, I articulate the normative attitudes to the cost of error implied by the location, shape, and curvature of an agent's epistemic risk function. In section 6, I develop the duality between risk and entropy. Section 7 extends the approach to more general sample spaces. Finally, section 8 explores the relationship between epistemic risk, the selection of priors, and the Laplacean POI.

2. Background. In financial analyses the expression 'value at risk' denotes the quantity (in monetary terms) that a firm or financial portfolio, say, stands to lose. As we seek to develop a theory of epistemic risk, we can begin by asking a related question: When an agent adopts a credence function that the theory deems risky, what is the *value* under risk? I can think of two approaches we might take to the epistemic value in jeopardy: the alethic approach, which I will develop, where the value is accuracy, and the modal approach, developed recently in Pritchard (2017), where the value is knowledge. The alethic approach is especially appropriate to Bayesian epistemology, since finely grained beliefs can approach the truth to various degrees, whereas the modal approach may be best suited for traditional full-belief approaches, in which justification plays a central role.

The Alethic Approach. I start with the veritist premise that the primary source of epistemic value associated with an agent's beliefs or credences is the extent to which they represent the world correctly (Goldman 1999, 2002). In the Bayesian context, veritism suggests that an agent should strive to adopt

https://doi.org/10.1086/703552 Published online by Cambridge University Press

high credences in truths and low credences in falsehoods. These are sometimes referred as the Jamesian goals (e.g., Pettigrew 2016b; Horowitz 2018), a reference to James (1896). But in identifying an appropriate credence function, we must strike a balance between these goals. To take an example from Levinstein (2017), while an agent could avoid massive inaccuracy by having credences close to 0.5, assuming an underlying accuracy measure that is symmetric, she would thereby sacrifice the epistemically valuable state of being highly accurate. Some agents may find this a valuable trade-off. If we suppose that suspension of belief is similar to credences close to 0.5, then Thomas Jefferson's well-known sentiment regarding ignorance is a good example. He writes, "ignorance is preferable to error, and he is less remote from the truth who believes nothing than he who believes what is wrong" (Jefferson 1785/1832, 46). Others may not. James himself notes that "a certain lightness of heart seems healthier than [such] excessive nervousness [about error]" (1896, 339).[2] Van Fraassen (1984) adopts a similar perspective.

On the view I develop, the way we strike this balance, given an underlying measure of accuracy, reflects our attitudes to epistemic risk. The epistemic risk function encodes this trade-off between confidently believing the true and confidently disbelieving the false. But the theory is more general than that. On larger sample spaces, the epistemic risk function reflects the way we balance approaching error in the direction of every possible outcome. In short: the normative value at risk is accuracy, and attitudes to risk of graded error are reflected by the epistemic risk function I will develop. This is similar to the relationship between risk and utility in ordinary decision theory, where the agent's attitude to ordinary risk is encoded in the curvature of her utility function.

The Modal Approach. Alternatively, we might think that what is at risk in epistemology is not the risk of error (i.e., the potential of holding a false belief or inaccurate credence) but rather the risk of holding a belief that, while true or accurate, fails to constitute knowledge. This idea emerges out of anti-luck approaches to epistemology, where safety is central to justification. For Pritchard (2007), an agent's belief is safe if it remains true in most nearby possible worlds in which the agent holds the belief in the same way as in the actual world. As a result, a belief's degree of risk is determined by the modal closeness of worlds in which the belief is similarly held but in fact false (Pritchard 2017).

Consider a simple case in which you are to formulate a belief about whether a lottery ticket will win. On Pritchard's account, a belief or high credence that the ticket will not win can count as risky, even though it is over-

2. In subsequent discussion, James appears to walk back his endorsement of lighthearted inquiry, at least in scientific pursuits.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

whelmingly likely to be accurate, because the worlds in which I am wrong (in which I win the lottery) are extremely similar to the actual world. Since risk is given in terms of a modal notion of closeness, rather than a measure-theoretic one, the description of risk is not necessarily sensitive to an agent's honest assessment of the probabilities involved. While this is not a problem if the underlying value at risk is knowledge, where justification is often understood in terms of safety or sensitivity, it does suggest that the modal account may be inappropriate for Bayesian epistemology.

Related to this, we could follow Buchak (2013) and attempt to construct a theory of epistemic risk where the risk attitude is given by a parameter that is independent of the agent's utility function. However, since Buchak's theory is a nonexpected utility theory, we would then carry the burden of explaining how it can be made compatible with the prevailing framework of epistemic utility that has emerged from, for example, Joyce (1998, 2009), Greaves and Wallace (2006), and Leitgeb and Pettigrew (2010). It is important on this framework, as it is in von Neumann and Morgenstern (1944) and Savage's (1954) ordinary expected utility theories, that failing to maximize expected utility is not rational. This is not true on Buchak's approach.[3]

3. Formal Framework. Following the literature, I adopt the useful fiction that an agent is able to choose between competing credence functions. Thus, credence functions will be the object of risk, and ultimately we seek to compare and rank them in terms of their riskiness. As suggested above, it is natural in this framework to suppose that one credence function is riskier than another if the agent stands to lose more in terms of accuracy or that variability in accuracy outcomes is greater. This resembles in some respects Peirce's (1879/1967) notion of the "economy of research."[4] I develop the theory care-

3. Since Buchak's risk-averse agents can rationally prefer one bet to another on the basis of outcomes that are identical between the two bets depending on how those outcomes affect the global distribution of the bet's payoffs, they can violate the independence axiom of expected utility theory. As a result, for such agents there is no probability measure and suitable utility function under which we can represent them as maximizing expected utility. While Buchak defends this approach in ordinary decision making, where certain cases of preference reversal appear to be intuitively rational (such as the preferences many people express regarding the sequence of bets that form the basis for the Allais Paradox), it is not clear whether (and if so, how) such a defense would extend to the epistemic context.

4. Peirce says, e.g., "The doctrine of economy, in general, treats of the relations between utility and cost. That branch of it which relates to research considers the relations between the utility and the cost of diminishing the probable error of our knowledge" (1879/1967, 643). As Rescher (1976) emphasizes, inductive logic is, in Peirce's view, crucially dependent on economic considerations and reasonable assessment of the risk of different types of error as well as the value of correct verdicts. Subsequently, Levi (1962, 1974), Maher (1990, 1993), and Fallis (2007) have suggested similar approaches to epistemic risk.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

fully below. The remainder of this section provides a directed introduction to the relevant formal concepts.

I assume that an epistemically rational agent should adopt as her credence function a probability distribution whose expected inaccuracy is at least as low as any alternative distribution she might adopt.[5] This is Joyce's (1998) norm of *gradational accuracy*. It captures the veritist spirit in a context of fine-grained subjective uncertainty. Minimizing expected inaccuracy plays a role in epistemic utility theory similar to what maximizing expected utility plays in ordinary decision theory. To measure inaccuracy, we use a *scoring rule*. This is a two-place function $s : \{0, 1\} \times [0, 1] \rightarrow \mathbb{R}$, denoted by $s_v(p(h))$, that measures the inaccuracy of the probability assigned to h when the true outcome is v, where $v = 1$ if h is true and 0 otherwise.

Three properties of scoring rules will be relevant to my argument: truth-directedness, continuity, and strict propriety. *Truth-directedness* implies that $s_1(p)$ is a decreasing function of p and $s_0(p)$ is an increasing function of p.[6] Thus, moving closer toward the actual truth-value cannot make an agent worse off. *Continuity* implies that s_1 and s_0 are continuous functions of p. This enables us to avoid arbitrarily small changes in credence leading to large changes in accuracy. Before we define strict propriety, we need to introduce one more concept. The expected inaccuracy of a probability distribution is the expectation of $s_v(p)$ evaluated with respect to the agent's beliefs, $b = b(h)$. In the binary case, this is

$$\mathrm{E}_b[s_v(p)] = bs_1(p) + (1 - b)s_0(p). \tag{1}$$

If this equation is (uniquely) minimized at $b = p$, the score is *(strictly) proper*. This means that a coherent agent can do no better in expectation, from the perspective of minimizing inaccuracy, than to adopt as her credence function the probability distribution that corresponds to her sincere degrees of belief.

One more property will be relevant to my argument. It is not presupposed in any of the theorems—rather, it will inform our discussion of the rationality of different attitudes to epistemic risk. We say that s_v is *0/1 symmetric* if, given two probabilities for h, $p(h)$ and $q(h)$, that are identical except that $p(h) = 1 - q(h)$, then $s_1(p(h)) = s_0(q(h))$.

I assume that an agent's normative attitudes to risk, if they are to be found anywhere, must be reflected in the prior the agent deems appropriate. As a result, in developing a measure of epistemic risk, we set aside for now considerations of updating and ask: regardless of one's evidence about a propo-

5. I restrict my attention to coherent agents for whom the credence function is a probability (this assumption can be relaxed). I generally define scoring rules in terms of *in*accuracy.

6. For compactness in simple binary cases, I often suppress the arguments of credence functions and write p instead of $p(h)$.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

sition, what structural features make one credence function riskier than another? Of course, it is also important to consider what makes one update riskier than another. Equivalently, how much epistemic risk might be justified by the agent's evidence? These are questions about dynamic epistemic risk, and I pursue them in subsequent work.[7]

4. Epistemic Risk: The Simple Case. Consider an agent formulating a credence $p(h)$ about a single proposition h. Regardless of h's content, we know that the agent's inaccuracy decreases as her credences get closer to the truth and that it increases as they get further away from it. Since s_1 is continuous and decreasing on [0, 1] with $s_1(1) = 0$, and s_0 is continuous and increasing on [0, 1] with $s_0(0) = 0$, the intermediate value theorem guarantees that there exists a point of intersection p^* for which $s_1(p^*) = s_0(p^*)$. For 0/1 symmetric scores, this is 0.5. For asymmetric scores it may be something else. Figure 1 illustrates this situation. Figure 1*a* depicts a symmetric score, whereas figure 1*b* depicts an asymmetric one.

The point p^* may be thought of as the least risky probability assignment in the following sense: if the agent's credence for h is given by p^*, her inaccuracy will be the same regardless of the actual truth-value for h. As a result, she knows with certainty how inaccurate she will be, even before she learns whether h is true or false.

It is natural to think of a guarantee in one's outcome as implying an absence of risk. Indeed, this is the purpose of ordinary insurance: to charge a premium for guaranteeing a particular outcome (and, in turn, removing risk)—hence, 'risk premium'. The outcome in insurance contexts is given in monetary terms. Here the same idea applies, but the relevant commodity is accuracy, and therefore the outcome is given in inaccuracy as measured by a scoring rule. Informally, therefore, we might identify p^* as the least risky probability in the sense that it guarantees a certain inaccuracy score, regardless of outcome. Since the choice of scale in constructing a risk measure is arbitrary, we may call p^* the risk-free credence and define it more formally as follows.

> **Risk-free credence.** Given a single proposition h the risk-free credence $p(h) = p^*$ satisfies the equation $s_1(p^*) = s_0(p^*)$.

Now suppose that the agent has a stronger credence for h, say 0.8. Then if h is true, her inaccuracy will be very low, but if h is false, her inaccuracy will be quite high. Since $p(h) = 0.8$ creates an opportunity for the agent

7. Note that every scoring rule may be associated with a measure of divergence between a prior and a posterior (Savage 1971). I use this link to connect the theory of risk developed here to a theory of dynamic risk for competing update rules.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

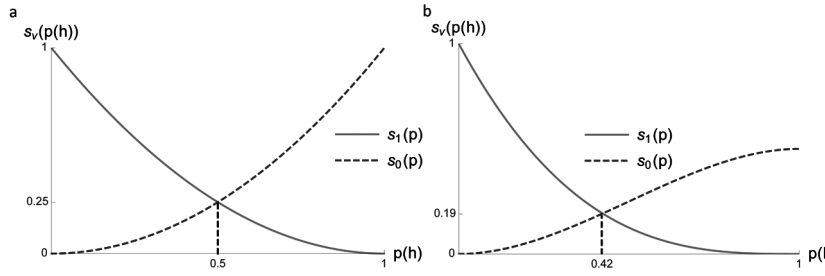


Figure 1. Risk-free probability p^*, where $s_1(p^*) = s_0(p^*)$: *a*, symmetric score; *b*, asymmetric score. Color version available as an online enhancement.

(the probability of doing better) together with a corresponding potential cost (the probability of doing worse), it is in this sense a riskier credence relative to p^* on the alethic approach. A natural measure for this increase in risk is the spread between s_1 and s_0, as depicted by the shaded areas in figure 2, because this quantity increases monotonically with shifts of probability to the tails of the risk-free distribution.[8] Figure 2*a* depicts the increase in risk from a $0/1$ symmetric score's risk-free credence, whereas figure 2*b* depicts the increase in risk from an asymmetric score's risk-free credence.

Notice, however, that it is less sensible to speak about one credence function being riskier than another if we vary the number of possible outcomes in the sample space. With three outcomes instead of two, the risk-free probability would occur where $s_1(p) = s_2(q) = s_3(1 - p - q)$. Assuming a $0/1$ symmetric score, this would be the uniform distribution $p = q = 1/3$. So to evaluate the riskiness of a credence function over three outcomes, we should measure the "spread" from the risk-free distribution for this larger space (we will see how to do this later). In light of these remarks, we may define a risk measure for the single proposition case as follows.

> **Epistemic risk.** Given a single proposition h and a risk-free credence p^*, the risk associated with investing credence $p < p^*$ in h is
>
> $$R(p) = \int_{p}^{p^*} |s_1(t) - s_0(t)|\, dt.$$
>
> For $p > p^*$, the bounds of integration are reversed. For $p = p^*$, $R(p) = 0$.

8. One may also consider the absolute value $|s_1 - s_0|$, as Joyce (2015) suggests. These two notions are closely related. I opt for density because it encodes more information about the agent's normative attitudes to risk, as it is sensitive to the curvature of the scores between the risk-free point and the target credences. As a result, this approach may be thought of as a more complete measure of a credence function's risk.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

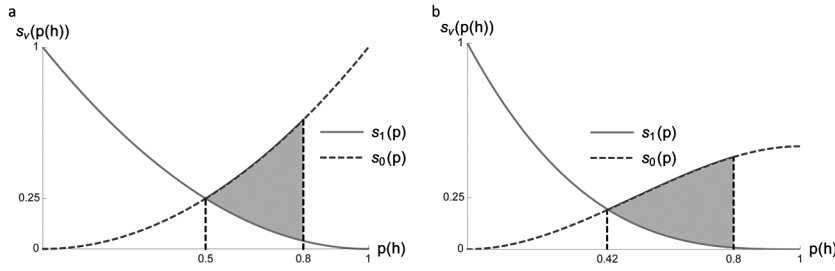


Figure 2. Increasing epistemic risk: *a*, symmetric measure of risk; *b*, asymmetric measure of risk. Color version available as an online enhancement.

Provided the scoring rule is continuous, the risk function will be likewise continuous. Its local maxima will occur at $p(h) = 0$ and $p(h) = 1$. Since the scoring rule must be monotonically decreasing as the credence approaches the true value, risk monotonically increases away from the risk-free credence.[9]

5. Risk and Normativity. Any move away from the risk-free credence threatens to increase inaccuracy, by either increasing confidence in h when it is false or decreasing confidence in h when it is true. Whether one deems the direction important reflects a substantial normative attitude toward the cost of approaching different types of error. As $p(h)$ goes up, one risks increasing inaccuracy in the direction of a false positive (type I) error. Meanwhile, as $p(h)$ goes down, one risks increasing inaccuracy in the direction of a false negative (type II) error.[10] It is doubtful that the only rational attitude to these types of error is indifference (as $0/1$ symmetry suggests). Being solely concerned with the truth, as Gibbard (2008) points out, does not commit one to a particular way of valuing accuracy. As a result, we want our measure of risk (and associated scoring rule) to reflect different trade-offs that agents might make between moving toward either type of error.

For example, h could be the outcome of a coin toss, where unit increases in inaccuracy in the direction of falsely predicting h (heads) are about as bad as unit increases in inaccuracy in the direction of falsely predicting its negation (tails). This set of attitudes to error is adequately captured by a $0/1$ sym-

9. I define epistemic risk with respect to the Lebesgue measure on the real line. It would be interesting to explore how the results below fare under different choices of measure.

10. I use the false positive/false negative distinction for illustrative purposes. The nomenclature can be misleading since we could redescribe the risk of increasing $p(h)$ as a risk of false negative error by deeming h to be the null hypothesis rather than its negation. What matters is the agent's relative attitude to approaching error in different directions, regardless of how we name them.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

metric score, such as the Brier score where $s_v(p) = (v - p)^2$, because an $\varepsilon > 0$ increase in inaccuracy in the direction of either s_1 or s_0 from any credence $k \in [0, 1]$ leads to a decrease in epistemic utility of $(k - \varepsilon)^2$. Figure 3*a* depicts this situation. As a result, the risk of $p(h) = 0.8$ (the shaded area to the right of the risk-free point) is equal to the risk of $p(h) = 0.2$ (the shaded area to its left). Indeed, they are reflections of each other around the risk-free point. Thus, an agent with this risk function is equally sensitive to unit increases in inaccuracy in the direction of either type of error.

Alternatively, h could be a very informative proposition that the agent is singularly pursuing. In this case, falsely believing h may be much better than falsely believing its negation. The latter may produce an enormous opportunity cost that delays or more permanently inhibits her search for the truth, for example, whereas the former may take the agent on a misleading line of inquiry that can be corrected through subsequent experimentation. In this example, unit increases in inaccuracy in the false negative error direction are worse than unit increases in inaccuracy in the false positive error direction.

One might worry that sensitivity to error appears to depend on considerations that are not purely epistemic. As a result, our measure of epistemic risk ultimately reflects these other values as well. But this is a feature of the account rather than a bug. It is compatible with the accuracy framework for an agent to have pragmatic reasons for the particular way in which she values accuracy. For example, it is reasonable for a weather forecaster to care more about false negative mistakes when the hypothesis is "there is a tornado nearby." This consideration can be a perfectly good reason for identifying a measure of inaccuracy.

Such attitudes to error are better captured by an asymmetric score whose risk function puts more weight on false negative increases in inaccuracy. An example of this is the score considered in Joyce (2009), where $s_1(p) = (1 - p)^3$ and $s_0(p) = (p^2/2)(3 - 2p)$. Like the Brier score, this score is strictly proper, continuous, and monotonic. But unlike the Brier score, an ε increase

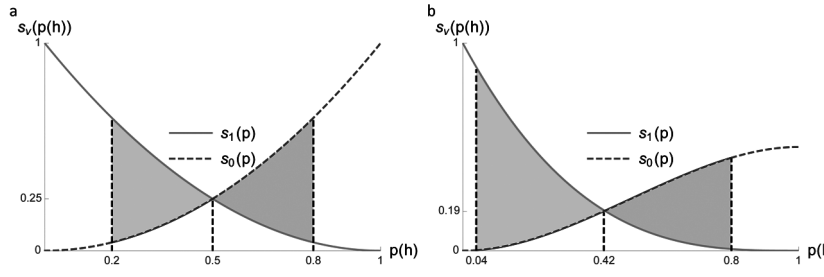


Figure 3. Epistemic risk as tolerance for different types of graded error: *a*, symmetric measure of risk; *b*, asymmetric measure of risk. Color version available as an online enhancement.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

in inaccuracy in the direction of s_1 from $p(h) = k$ leads to a decrease in the epistemic utility of $(k - \varepsilon)^3$, whereas an increase in inaccuracy in the direction of s_0 leads to a decrease in the epistemic utility of $\varepsilon^2(3 - 2\varepsilon)$. This situation is depicted in figure 3*b*. For this score, a unit move away from the risk-free credence in the direction of a false positive error leads to a smaller increase in risk (the shaded area to the right) than a correspondingly large move away from the risk-free credence in the direction of a false negative error (the shaded area to the left). As a result, the risk of $p(h) = 0.8$ is not equal to the risk of $p(h) = .04$ (nor for that matter is it equal to $p(h) = 0.2$).[11]

The symmetry of the embedded scoring rule is encoded in the risk function itself. In particular, it is reflected by the location of the risk function's minimum. As figure 4 shows, a risk function associated with a $0/1$ symmetric score will reach its minimum at $p(h) = 0.5$ (fig. 4*a*), whereas if the risk reaches its minimum elsewhere on the unit interval the embedded score must be asymmetric (fig. 4*b*).

I refer to risk functions such as the one in figure 4*a* as *symmetric*: it reaches its minimum at $p(h) = 0.5$, and its shape on $[0, 0.5)$ is a reflection of its shape on $(0.5, 1]$. Symmetry in the risk function is related to the $0/1$ symmetry of the scoring rule: a scoring rule is $0/1$ symmetric only if its associated risk function is symmetric.

Therefore, we should distinguish at least two different ways of valuing accuracy: a symmetric risk function corresponds to a way of valuing accuracy in which moving away from the truth in either direction is equally bad, whereas an asymmetric risk function implies a way of valuing accuracy in which unit changes in the direction of false positives/negatives get weighted differently at different credal values. Indeed, they may not be weighted equally at any place. It is not enough, therefore, to declare that we should seek truth and avoid error. Such an epistemic norm is underspecified. We need to decide further how to trade-off the potential costs of different types of mistakes. The epistemic risk function is flexible enough to encode different ways of balancing the competing costs.

So far we have exploited only the location of the risk function. But the risk function in figure 4*b* is not just shifted to the left. Speaking picturesquely, it is also pressed against the *y*-axis. As a result, there is both a within and between difference in its *concavity*: it is (*a*) steeper to the left of its risk-free

11. Note that there will be an equally risky point in the direction of a false negative mistake as $p(h) = 0.8$, i.e., the point $p(h) = \gamma$, where $\int_{\gamma}^{.42}(s_1 - s_0)\,dt = \int_{.42}^{0.8}(s_0 - s_1)\,dt$. But since this particular score is relatively more sensitive to moving in the direction of a false negative error, γ will be closer in probability to the risk-free credence than 0.8 is to the risk-free credence. Therefore, while permuting probabilities for symmetric scores does not affect their risk, for asymmetric scores permuting probabilities does not preserve risk. However, there exist isomorphisms that would preserve it.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

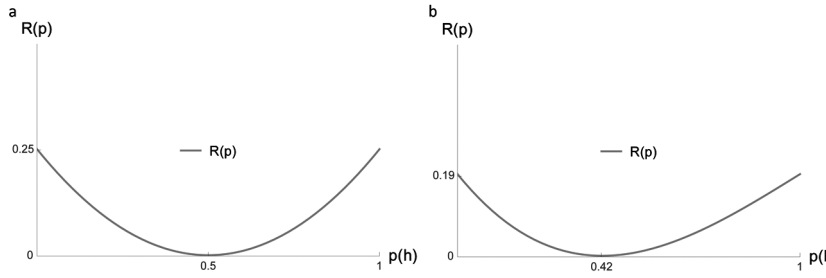


Figure 4. Epistemic risk function: *a*, symmetric risk function; *b*, asymmetric risk function. Color version available as an online enhancement.

point than it is to its right, and (*b*) it is not equally concave as compared to the risk function in figure 4*a*, whose embedded score is symmetric. These properties add further texture to the proposed measure of risk, reinforcing the idea that risk is a measure of alethic sensitivity to error. To exploit the concavity of the risk function, we need to revisit another quantity.

Let $h(p) = s_1(p) - s_0(p)$. For example, when $p = 0.8$, $h(p)$ is a measure of the length of the dashed vertical line connecting s_1 and s_0 at 0.8 in figure 3. The antiderivative of $h(p)$ is $R(p)$. As a result, our definition of epistemic risk implies that $R'(p)$ is equal in absolute value to $h(p)$. This means that the rate at which risk increases as we move away from the risk-free point reflects the increase, in absolute value, between the agent's best and worst outcomes. As a result, while the risk function itself reflects the agent's relative sensitivity to unit increases in inaccuracy in the direction of different types of error, its first derivative reflects, instead, the agent's local sensitivity to risk as a function of her current credence. It is a measure of marginal increases/decreases in risk. For example, the derivative of the risk associated with the Brier score is $2p - 1$. As a result, marginal changes in credence away from the risk-free point lead to a constant increase in risk, as figure 5*a* shows. If we let ΔFP stand for marginal increases in false positive inaccuracy and ΔFN stand for marginal increases in false negative inaccuracy, then a symmetric risk function (such as the Brier score's) implies that ΔFP = ΔFN.

By comparison, the derivative of the risk associated with the asymmetric score we have been considering is $-(3/2)p^2 + 3p - 1$ (fig. 5*b*). For this score, marginal changes in credence away from the risk-free point in the direction of a false negative error lead to bigger changes in risk relative to marginal changes in credence away from the risk-free point in the direction of a false positive error. The agent applying this particular asymmetric score is more worried about marginal increases in false negative inaccuracy than she is about marginal increases in false positive inaccuracy. For this particular asymmetric risk function, ΔFN > ΔFP. This corresponds to the exam-

https://doi.org/10.1086/703552 Published online by Cambridge University Press

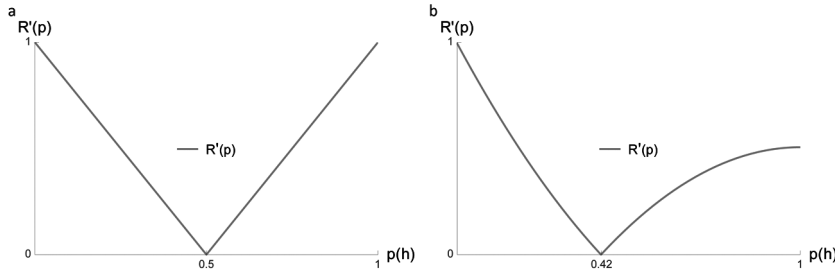


Figure 5. Rate of change in epistemic risk: *a*, constantly increasing epistemic risk aversion; *b*, unequally increasing epistemic risk aversion. Color version available as an online enhancement.

ple described above, where h is so important that rejecting it leads to substantial epistemic opportunity cost.

Moreover, marginal increases in risk taper off as the agent approaches categorical false positive error. This makes sense from a Bayesian perspective of scientific inquiry, since having credence .05 in a true and important proposition is not that different from having credence .01 in the same proposition. In both cases, the agent will likely not pursue the idea further. Meanwhile, given her concern about false negative error, her anxiety in that direction persists, leading to near constant marginal changes in risk across the whole [0, .42) subinterval.

We can see this dimension of the agent's attitude to risk in the second derivative of the risk function. Gibbard (2008, 9) calls $R''(p)$ an indicator of the urgency the believer ascribes to getting credences right, by her lights, in the vicinity of p. For the Brier score, $R''(p) = 2$. No matter where the agent's credence is on the unit interval, her local sensitivity to being mistaken remains the same. For our asymmetric score, $R''(p) = 3 - 3p$. This is exactly what we described in the previous paragraph. This is a constantly decreasing function from 0 to 1. The agent's peak local sensitivity to error occurs at categorical false negative error and slowly tapers off as she approaches false positive error. Given the sensitivity of this particular score to false negatives, that is to be expected because $p(h) = 1$ is where false negatives are eliminated altogether.

One might wonder whether this is a reasonable attitude to false positive error. But this example should not be taken as an endorsement of this particular risk function. Rather, I use it to illustrate the flexibility of the proposed approach to capturing a wide range of attitudes to epistemic risk. The concavity of the risk function resembles in some respects the Arrow/Pratt measure of risk aversion for ordinary economic prospects, where the normalized

https://doi.org/10.1086/703552 Published online by Cambridge University Press

second derivative reflects an agent's relative sensitivity to the ordinary risk of monetary loss (Pratt 1964; Arrow 1965, 1971).[12]

6. Risk and Generalized Entropy. When equation (1) is uniquely minimized at $b = p$ (i.e., the scoring rule is strictly proper) it may be rewritten as follows:

$$\mathrm{E}_p[s_v(p)] = ps_1(p) + (1 - p)s_0(p). \tag{2}$$

Following Grunwald and Dawid (2004), I refer to this function, $E_b[s_v(p)]$ in which $b = p$, as $H(p)$, the *generalized entropy*. Let me explain why, as this will be relevant later. Suppose $w(p)$ is a measure of information conveyed by learning that the event h occurs with probability p. What conditions should w satisfy? This is the question Shannon (1948) seeks to answer. His famous result is a representation theorem showing that the logarithmic construction $w(p) = k \log(p)$ uniquely satisfies several intuitively plausible constraints on a measure of information—namely, that w should be a decreasing, continuous, and additive function of p. By the same token $-w(p)$ measures a lack of information, and Shannon entropy is the expectation of $w(p)$ with $k = -1$.

In the binary case, Shannon entropy becomes $-[p \log(p) + (1 - p) \log(1 - \mathrm{p})]$. This is equivalent to the expected inaccuracy of the log score, which is strictly proper. But we can think more generally about an entropy function H associated with other strictly proper scoring rules—the weighted average of a different strictly proper score function of the probability. This is generalized entropy. Generalized entropy is an important building block in epistemic utility theory because Savage (1971) gives us a recipe for deriving strictly proper scores from entropy by showing that every twice differentiable concave entropy function corresponds to a strictly proper scoring rule, as follows:

$$s_v(p) = H(p) + (v - p)H'(p), \tag{3}$$

where v is the $0/1$ truth-value for the event in question. This relationship is extremely useful. As long as we start from a twice differentiable $H(p)$ concave on $[0, 1]$, we can derive a continuous, truth-directed, strictly proper score.

The entropy function H is closely related to our measure of epistemic risk, R. For example, for the Brier score, risk is equal to $p^* - p(1 - p)$, whereas

12. It has been noted in the literature that the convexity of a scoring rule implies aversion to epistemic risk in the following sense: suppose an agent is offered a pill that would, with equal probability, raise or lower her credence in h by $k \in [0, 1]$. If the scoring rule is convex, such a pill would look unattractive in expectation because losses are weighted more heavily than gains (Joyce 2009).

https://doi.org/10.1086/703552 Published online by Cambridge University Press

entropy is $p(1 - p)$. This relationship is depicted in figure 6*a*. Meanwhile, for our asymmetric score risk is $p^* - p(p - 1)(p - 2)$, whereas entropy is $p(p - 1)(p - 2)$. We can see this in figure 6*b*.

The following theorem establishes that this duality between generalized entropy and epistemic risk holds for all strictly proper scoring rules.

> **Theorem 1.** For strictly concave and twice differentiable entropy function H and risk function R defined on [0, 1], $R(p) + H(p) = k$, where $k = \min_p R(p) = \max_p H(p)$.
>
> *Proof.* See the appendix.

In other words, the sum of risk and entropy is constant: risk + entropy = k. In general, therefore, entropy is a scaled reflection of epistemic risk around the risk-free point $R(p^*) = k$, as figure 6 suggests. But the risk-free credence is also the maximum entropy credence. Therefore, rearranging the duality equation suggests that epistemic risk may be expressed as a measure of entropic change from the maximum entropy credence to the target credence: $R(p) = H(p^*) - H(p)$. We will use this expression below to develop a more general measure of epistemic risk.

Note that since we have defined risk in terms of expected inaccuracy, it follows that for a fixed credal value $p(h) = k$, the risk associated with k is constant. In other words, given any credence, while it is true that the agent's accuracy for that credence is a random quantity, because the agent does not know whether h is in fact true or false, the amount of risk the agent assumes is fixed, because it is a function of the distance between those two outcomes. As a result, we cannot evaluate epistemic risk from a different credal point. Therefore, while we may consider accuracy in expectation, on this account we should avoid talking about expected epistemic risk. This will be relevant as we consider the import of epistemic risk to the selection of priors in section 8.

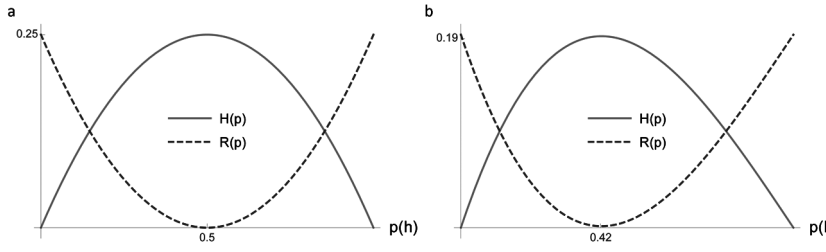


Figure 6. Duality between epistemic risk and entropy; *a*, symmetric case; *b*, asymmetric case. Color version available as an online enhancement.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Since epistemic risk is dual to entropy, one might question whether we need to introduce a notion of risk, given the large literature on entropic inference.[13] Rather than speaking in terms of increases in epistemic risk, we could instead describe the same changes in terms of decreases in entropy. Although this is true for strictly proper scoring rules, with the effect that risk and entropy are often coextensive, they are independently motivated. We saw this while developing the notion of epistemic risk in terms of sensitivity to different types of graded error. That is, I am not arguing that the risk-free credence function is risk-free because it maximizes entropy. Rather, it is risk free, as we saw, because it eliminates variability in terms of the epistemic outcome. Strictly proper scoring rules have the feature that these two properties do not come apart. For many other scoring rules, we could eliminate variability without maximizing entropy. In such cases, the duality would not apply, and we could not measure epistemic risk in terms of entropic change.

Therefore, even though risk and entropy are extensionally equivalent for strictly proper scoring rules, thinking in terms of risk minimization is conceptually very different from thinking in terms of entropy maximization. An agent might prefer risk-free credences not because they do not go beyond the evidence, even though that might be true, but because from her perspective they give her the best balance of graded error costs. There is a conceptual difference between thinking in terms of minimizing the amount of information an agents brings into the inference problem (the entropic interpretation) and identifying an appropriate trade-off between different types of potential mistakes (the risk interpretation). As a result, we should not think of one concept being reducible to the other. The duality theorem shows that for many scoring rules, entropy and risk are two different ways of conceptualizing the same underlying epistemic facts.

Indeed, insofar as proponents of entropic methods reference risk, it is assumed that a credence function is risk averse because it maximizes Shannon entropy. Jaynes is the most ardent proponent of this position. For Jaynes, the maximum entropy distribution is the most conservative distribution in the sense that it does not permit us to draw any evidentially unwarranted conclusions because it is "as smooth and spread out as possible" subject to the data (Jaynes 1963/1983, 186). But consider an entropy function that reaches its maximum at $p(h) = 0.9$. An entropy-maximizing agent with this function would not be conservative at all in Jaynes's sense. In the absence of any data, she would predict h's occurrence with high confidence. Therefore, for asym-

13. For example, Jaynes (1957, 2003) defends maximum entropy methods for identifying priors, whereas Williamson (2010) goes further and defends updating by maximizing entropy as well. Seidenfeld (1986) contains a thorough discussion of the relationship between Bayesian epistemology and entropic methods.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

metric risk functions the least risky distribution will not be maximally uniform.

7. Epistemic Risk: The General Case. So far we have considered credence functions for a single proposition h. Now let the sequence $\{h_i\}_{i=1}^{n}$ form a partition on sample space S. The risk-free credence function becomes the distribution that solves the equation $s_v(p_i) = s_w(p_j)$ for all i, j and indicators of truth-value v, w. Since this expression is unwieldy with many outcomes, we can instead identify this as the point of maximum general entropy. Because entropy is the expected inaccuracy of a strictly proper scoring rule, expressing risk in terms of entropic change enables us to harness helpful properties of expectation.

To make use of these properties, we require a random variable and its cumulative distribution function (cdf). A cdf is just a different way of expressing a probability distribution. Let $X : S \to \mathbb{R}$ be a random variable that maps outcomes in the sample space to the real numbers and whose mass/density is given by $f(X = x)$. For each value of x the cdf, defined as $F(X \leq x) = \Sigma_{x_i \leq x} f(x)$ (for discrete X) and $\int_{-\infty}^{x} f(t)\,dt$ (for continuous X), gives us the probability that X is less than or equal to that value. For example, if the random quantity X represents the numerical outcome of a single toss of a die, then $F(X \leq 3) = 1/2$ and $F(X \leq 4) = 2/3$.

For our purposes, every outcome may be described in terms of the agent's inaccuracy if that outcome occurs. Therefore, we can define outcomes in terms of random variables as follows: let X be a random variable that maps outcomes from the sample space to the real numbers, where the real numbers represent inaccuracy given by s. For every valid probability distribution on the sample space, there exists an induced probability distribution on X that is likewise valid. The possible values of the random variable now represent inaccuracy scores. Many scoring rules will take values on a small subinterval of $\mathbb{R}$. For example, under the Brier score all outcomes are mapped to [0, 1]. Changing the underlying scoring rule will rescale the random variable. Therefore, when evaluating credence functions in terms of their epistemic risk, we need to identify a random variable that describes outcomes in terms of some particular measure of inaccuracy. With this in mind, we can define the risk-free cdf as follows.

> **Risk-free cdf.** Let $W \subseteq R$ be the image of scoring rule s. Given a random variable mapping outcomes from the sample space S to inaccuracy given by s, $X : S \to W$, the risk free cdf $P^* = \arg\max_P H_P(X)$.

To simplify, I denote the entropy of cdf P as $H(P)$ instead of $H_P(X)$ (a common abuse of notation, since entropy is a property of the probabilities). As emphasized above, P^* is not risk free because it maximizes entropy.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Rather, this is the probability assignment that eliminates variability in terms of epistemic outcome, which is how we defined the risk-free credence in the simple case. We can now extend our definition of epistemic risk as follows.

> **General epistemic risk.** Given a random variable $X : S \to W$, where W is defined as above, let cdf $P^* = \arg\max_P H(P)$. Then theorem 1 suggests that a natural extension of the notion of epistemic risk to larger partitions would be to define the epistemic risk of another cdf P by $R(P) = H(P^*) - H(P)$.

Recall that in the simple case, this definition was motivated as a measure of the "spread" between the agent's inaccuracy if the proposition is true and her inaccuracy if the proposition is false. It remains to be shown that the general definition given here is motivated by the same underlying conceptual framework.

To see that this is indeed the case, I draw on Rothschild and Stiglitz's (1970) notion of a mean-preserving spread. Informally, one probability distribution is a mean-preserving spread of another if the second is a transformation of the first obtained by pushing probability mass/density to the tails of the distribution without affecting its expected payoff. In the case of ordinary economic lotteries, distributions are given in terms of wealth. For example, a lottery that pays \$0 or \$10 with equal probability is a mean-preserving spread of one that pays \$6 or \$4, or one that guarantees \$5.

In the epistemic context, the outcomes of a "lottery" cannot be specified exogenously. Rather, the scale (i.e., scoring rule) is exogenous, but the outcome, given in terms of that scoring rule's inaccuracy, depends on the probability assignment itself. For example, assuming the Brier score, a credence $p(h) = 0.8$ in a single proposition h is effectively an epistemic lottery that pays $(1 - 0.8)^2 = .04$ if h is true and $(0 - 0.8)^2 = .64$ if h is false. Now consider a more extreme credence like $p(h) = 0.9$. The latter is a probabilistic spread of the former because it is a transformation accomplished by taking the probability assigned to h and making it even more extreme while at the same time taking the probability assigned to its negation and making it correspondingly more extreme in the opposite direction. Assuming the agent is coherent, there is a quantity that is preserved every time we spread out probability like this, namely, the simple mean given by $1/|S|$, where $|S|$ is the length of the partition. As long as we keep this quantity fixed, every such spread guarantees an increase in risk. In this sense, a mean-preserving spread of a credence function implies an increase in that credence function's epistemic risk. By expressing a credence function in terms of its cdf, we can give a general definition of mean-preserving spreads and prove this relationship.

For example, suppose $\{h_1, h_2, h_3\}$ is a partition on S and we want to measure the epistemic risk of credence function p (or equivalently, its cdf P)

https://doi.org/10.1086/703552 Published online by Cambridge University Press

given by $\langle 1/5, 3/5, 1/5\rangle$ under the Brier score.[14] Since the Brier score is 0/1 symmetric, we know that its risk-free credence function p^* is the uniform $\langle 1/3, 1/3, 1/3\rangle$.

Before we move on, note that everything we say below will hold for nonsymmetric scoring rules as well. To illustrate, we could use instead a nonsymmetric weighted quadratic score that determines the accuracy assigned to proposition h by $(v - pr(h))^2 + m/n$, where h is the mth cell in a partition and n is the number of cells. If ordering the cells is not appropriate, we could determine the weights some other way. The important thing is that the weights capture our attitudes to error with respect to each possible outcome. With three propositions again, the risk function of such a score would be minimized with the probabilities $\langle .17, .30, .53\rangle$. These are now the (nonuniform) risk-free credences. Notice that the probabilities increase from the first cell to the last. This is to be expected because such a scoring rule penalizes errors with increasing severity as we move up the sequence generating the partition. Because we are most sensitive with respect to errors in the direction of h_3, that outcome is most sticky, so to speak, and the risk-free credences are extra cautious in its direction.

In any case, to keep things simple, I will continue with the symmetric example. To evaluate the spread of our target credence function from the risk-free credences, we write the cdfs of both credence functions, P and P^*, as follows:

$$P^* = \begin{cases} 0 & \text{for } x < (1/3)^2 \\ 2/3 & \text{for } (1/3)^2 \le x < (2/3)^2 \\ 1 & \text{for } x \ge (2/3)^2 \end{cases} \qquad P = \begin{cases} 0 & \text{for } x < (1/5)^2 \\ 4/5 & \text{for } (1/5)^2 \le x < (4/5)^2 \\ 1 & \text{for } x \ge (4/5)^2 \end{cases}$$

Figure 7*a* depicts the plot of each cdf. The arrows indicate the spread in probability generated by moving from P^* to P. This is harder to visualize for discrete cdfs. To make the idea more intuitive, figure 7*b* depicts two arbitrary cdfs of a continuous random variable X, where one is a mean-preserving spread of the other.

Notice that for any value of X, representing an outcome in terms of inaccuracy, the area underneath the dashed (risky) curve is greater than or equal to the area underneath the solid (safe) curve. Following Rothschild and Stiglitz (1970), we can use this quantity to define mean-preserving epistemic spreads.

14. In general, I use lowercase for the mass/density and uppercase for the cdf.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

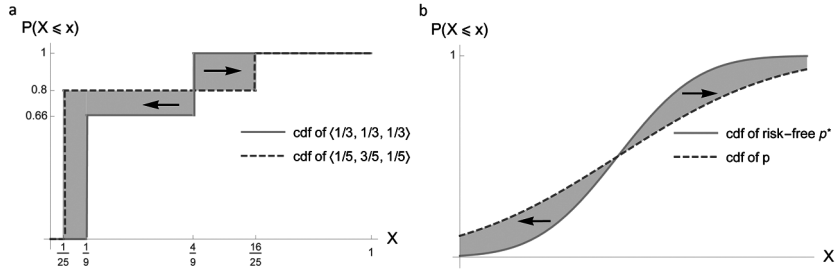


Figure 7. Mean-preserving epistemic spreads; *a*, discrete credence function; *b*, continuous credence function. Color version available as an online enhancement.

Mean-preserving epistemic spread. Given a random variable $X : S \to W$, where W is defined as before, let P and Q be two cdfs. Then Q is a mean-preserving epistemic spread of P if, for all x, $\Sigma_{i=0}^{x} P(t_i) \leq \Sigma_{i=0}^{x} Q(t_i)$ (if X is discrete) and $\int_0^x P(t)\,dt \leq \int_0^x Q(t)\,dt$ (if X is continuous).

In the single proposition case, this implies that one probability $q(h)$ is a mean-preserving epistemic spread of another probability $p(h)$ if $|s_1(p) - s_0(p)| < |s_1(q) - s_0(q)|$. This is consistent with our definition of epistemic risk in the simple case as the integral of the absolute difference between s_1 and s_0. Therefore, by using mean-preserving epistemic spreads to measure risk, we measure the difference in area underneath the risk-free cdf and the target cdf. In figure 8, this is the difference of the two rectangles labeled *A* and the rectangle labeled *B*.

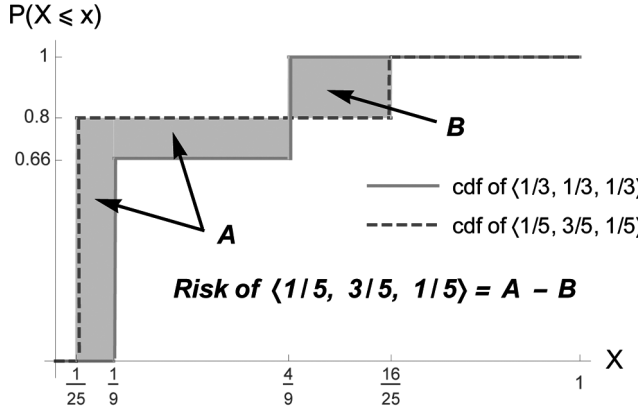


Figure 8. Epistemic risk as entropic change. Color version available as an online enhancement.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

This measure of epistemic risk, in terms of the change in area underneath the cdf, developed by analogy to Rothschild and Stiglitz's (1970) approach to ordinary risk, preserves the motivation given for measuring epistemic risk in the simple case as sensitivity to approaching different types of error. In the general case, however, epistemic risk reflects an agent's sensitivity to graded inaccuracy with respect to any given outcome in the sample space. As a result, we no longer have type I and type II errors only. Instead, we have n error types for $|S| = n$ possible outcomes.

We are now in a position to show that our definition of epistemic risk in terms of entropic change corresponds to the general interpretation of epistemic risk given in terms of mean-preserving epistemic spreads. For any given cdf P, as the area underneath it, given by $\Sigma_{i=1}^{n} P(x_i)$ (for discrete X) or $\int_X P(x)\,dx$ (for continuous X), decreases, the quantity $1 - \Sigma_{i=1}^{n} P(x_i)$ (for discrete X) or $1 - \int_X P(x)\,dx$ (for continuous X), increases. In figure 8, for example, for each cdf, this is the area to its left and bounded above by the line $P(X \leq x) = 1$. This quantity is equal to the expectation of X. This relationship is a consequence of Fubini's Theorem. Importantly for us, since X maps outcomes to inaccuracy scores, the expectation of a random variable X with cdf P is precisely the entropy of P, $H(P)$, provided the underlying inaccuracy scale given by s is strictly proper. Furthermore, on any given sample space S, the risk-free cdf will be the cdf that has the smallest area underneath it. Equivalently, it will be the cdf that has the largest area to its left. We can see this in figure 8. In more familiar words, the risk-free credence is the maximum entropy credence. Again, however, it is risk free not because it maximizes entropy but rather because this is the point where the agent's sensitivity to graded error in the direction of every possible outcome in the sample space is equal. And again it turns out, as in the simple case, that for strictly proper scores this is also the point that maximizes entropy. Therefore, as measured in terms of mean-preserving epistemic spreads, risk may be given as the difference between the entropy of the risk-free cdf and the target cdf. This is precisely the quantity *A–B* in figure 8, and it corresponds exactly to how we have defined epistemic risk, as $H(P^*) - H(P)$.

For example, consider the cdfs depicted in figure 8. To measure the risk of P, we first determine the entropy of the risk-free P^*. The area to the left of its cdf is a sum of two rectangles: one of length 1/3 and width 2(1/3) and another of length 2(1/3) and width 1/3. This is 4/9. Next, we determine the entropy of P. Following the same approach, we get 8/25. Since risk is given in terms of entropic change, the risk of P is $4/9 - 8/25 = .12$. This leads to the following theorem.

> **Theorem 2.** Given a random variable $X : S \rightarrow W$, where the underlying scoring rule s is strictly proper, and two cdfs P and Q, if P is a mean-preserving epistemic spread of Q, then $R(P) > R(Q)$.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Proof. See the appendix.

As a result, every mean-preserving epistemic spread increases variability in the underlying outcomes, increases risk, and (if s is strictly proper) decreases entropy.

Since the approach we have developed requires identifying an inaccuracy scale before evaluating the risk of a credence function, one might reasonably wonder how general the risk ordering of credence functions will be. For example, suppose we have the same two credence functions as in the previous paragraph, $p^* = \langle 1/3, 1/3, 1/3 \rangle$ and $p = \langle 1/5, 3/5, 1/5 \rangle$, but we define epistemic outcomes logarithmically instead of quadratically. That is, the x-axis now measures inaccuracy in terms of the log score. The y-axis still measures cumulative probability. Would it still be the case that $R(P) > R(P^*)$? If so, would the risk order be preserved for any arbitrarily chosen set of cdfs?

For most families of scoring rules considered in the literature, including some improper scores, the risk ranking of credence functions will be consistent. This includes the Brier, log, spherical, and absolute value scores. But it does not include the asymmetric score we have been considering throughout. This is because the asymmetric score has a different risk-free point, and risk is measured in terms of deviation from that risk-free point. Of course, if we take two asymmetric scores with the same risk-free point, wherever it happens to be, then it is very likely that the risk ordering between them will be consistent. Specifically, for any two scoring rules, if they share the same risk-free point, and their risk function is convex, then the risk order of credence functions between them will be consistent. The following theorem captures this relationship.

> **Theorem 3.** Given a random variable $X : S \to W$, where $W \subseteq \mathbb{R}$ is the image of scoring rule s, let $V = \{P_1, \ldots, P_n\}$ be a set of cdfs for X. Given a random variable $Y : S \to W^*$, where $W^* \subseteq \mathbb{R}$ is the image of scoring rule s^*, let $U = \{Q_1, \ldots, Q_n\}$ be a set of corresponding cdfs for Y. This means that for each outcome $h \in S$, the probability assigned to h by P_i is equal to the probability assigned to h by Q_i, but whereas in the first case the outcome h is described by s, in the second case it is described by s^*. Suppose (1) s and s^* are truth-directed scoring rules, whose risk functions R and R^* are such that (2) $R'' > 0$, $R^{*\prime\prime} > 0$, and (3) $\arg\min R = \arg\min R^*$. Then $R(P_i) > R(P_j)$ if and only if $R(Q_i) > R(Q_j)$.
>
> *Proof.* See the appendix.

This result expands the reach of our approach to epistemic risk to the vast majority of commonly considered families of scoring rules.

That is not to say, however, that all information encoded in the risk function will be preserved across different scoring rule transformations of it. Con-

https://doi.org/10.1086/703552 Published online by Cambridge University Press

sider the Brier and log risk functions. While they are both convex and share the same risk-free point, their derivatives are different. As a result, while a Brier-to-log transformation preserves an agent's risk ordering, it does not preserve her attitudes to unit changes in inaccuracy nor does it preserve her local sensitivity to marginal changes in risk. We could have two agents who rank two prospective credence functions equally in terms of risk, yet while one agent finds that degree of risk tolerable, the other considers it to be inappropriate, because of differences in the way they evaluate the potential cost of increasing graded inaccuracy in the direction of any given outcome. This is to be expected, however. We would not want a risk function that erases well-known differences between these scores. As Selten (1998) emphasizes, the log score is hypersensitive in the sense that one's inaccuracy goes to infinity as the probability assignment goes to 0 or 1. This hypersensitivity is reflected in the curvature of its associated risk function.

Before we move on, it is worth pausing to clarify the relationship between measures of epistemic risk and measures of attitudes to it. In ordinary economic theory, mean-preserving spreads are used to generate a partial ordering of stochastic alternatives in terms of their degree of risk. Meanwhile, the curvature of an agent's utility function reflects her sensitivity to risk. Rothschild and Stiglitz (1970) are so influential because they show that risk-averse agents prefer less risky lotteries to more risky ones, as we would expect. In the epistemic case, we have a similar relationship, to an extent. The epistemic risk function enables us to rank prospective credence functions in terms of their risk. Meanwhile, an agent is risk averse if her scoring rule is convex. And in the absence of information, an agent with a convex scoring rule would prefer a less risky credence function to a more risky one.

Risk increases with mean-preserving spreads in accuracy, and all truth-directed scoring rules with the same risk-free point agree on this ranking, regardless of their convexity. But the way risk is judged by agents to increase—the rate and acceleration of the increase in risk—reflects the agents' attitudes to risk. Such attitudes originate in the curvature of their scoring rule (i.e., the way they value accuracy). As a result, the risk ranking of credence functions in epistemology is not completely independent from the agents' attitudes to risk in the way that stochastic dominance is independent of a utility measure. This is inescapable, however. In economic theory, the probabilities and outcomes are both exogenously specified (e.g., the monetary prizes are determined in advance and their associated probabilities given by a roulette wheel), whereas in epistemology the accuracy outcomes are a function, in part, of the way probabilities are distributed. In other words, in the epistemic case one's payoff is directly determined by the probability one assigns to that outcome. By contrast to economic lotteries, we do not have probabilities for the outcomes that are specified from the outside and independent of the "money" (i.e., probability) wagered on them.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

8. Risk, Priors, and the Principle of Indifference. By developing a theory of accuracy dominance Joyce (1998, 2009) gives us a powerful tool for evaluating the quality of an agent's beliefs. The theory of epistemic risk enables us to go further in terms of our understanding of the normative dimensions of an agent's credal state. One might ask how these attitudes to risk will manifest themselves. Nearly everyone in the literature agrees that an agent should choose the credence function that, in light of her evidence, minimizes her expected inaccuracy. As a result, attitudes to risk are not going to play a direct role in one's choice (fictional or otherwise) of what to believe. But this is not the role of risk in ordinary expected utility theory either. We do not consult our sensitivity to risk in order to make a choice. Instead, our choice reflects our attitudes to risk. Roughly the same is true in the epistemic case.

However, risk attitudes can play a more direct role at the beginning of one's epistemic practices: in particular, agents' attitudes to risk (i.e., how much of it they are willing to assume) together with the shape of their risk function (e.g., symmetric, nonsymmetric) can motivate a choice of prior in the absence of information. The Laplacean POI is often given as a crude guide for this task. In the absence of information to privilege one outcome over others given a partition of the sample space, one should assign equal probability to each. It is assumed, therefore, that given an appropriate partition the POI recommends uniform credences. The most well-known problems with this principle stem from its association with uniformity. In particular, the uniform distribution over one partition may be logically inconsistent with the uniform distribution over a simple transformation of that partition.[15] Epistemic risk provides a new perspective on the POI—one that enables us to dissociate it from uniformity.

If we have an agent whose risk function is convex and symmetric, then the credence function that obtains minimum epistemic risk will be uniform. This is because a symmetric and convex epistemic risk function is associated with a 0/1 symmetric scoring rule. The proof of this is trivial. Under such a score, an agent would be indifferent between taking a bet whose payoffs are given in terms of inaccuracy on any proposition when the probabilities assigned to them are equal. Therefore, minimizing epistemic risk under these conditions suggests the same credence function as the ordinary Laplacean POI. As a result, we can think about the Laplacean POI as a heuristic for identifying the risk-minimizing credences under a very particular class of risk functions. However, it is only in the special case of convex and symmetric risk functions that indifference and uniformity are guaranteed to coincide.

15. For example, as John Venn first observed, a uniform distribution over X is not uniform over X^2. Van Fraassen (1989) makes this point vividly using the example of a box whose dimensions are unknown and may be measured in terms of side length or volume.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

This line of thought enables us to go further. We can consider risk-minimizing heuristics for risk functions that do not satisfy these conditions. For such risk functions, the credences that obtain minimum risk will be such that the agent is indifferent with respect to taking a bet on any outcome, but they will not be uniform. Therefore, we may think of each risk function as having its own associated POI, but when convexity and symmetry are not satisfied it is not guaranteed to be Laplacean. By recasting the Laplacean POI as a risk minimization principle, we can identify the normative commitment presupposed in its endorsement. In particular, it requires the agent to care equally about approaching different types of error. Just as importantly, we can sever its association with uniformity. For scoring rules that are not 0/1 symmetric, there will be a credence function in which the agent is indifferent regarding the outcomes, but it will not be uniform.[16]

To illustrate the relationship between epistemic risk and indifference principles, suppose we have two agents, A and B, whose risk functions, given a simple partition involving h and its negation, are given by the symmetric and asymmetric risk functions we have been considering, as follows:

$$r_A(p) = p^* - p(1 - p) \qquad r_B(p) = p^* - p(p - 1)(p - 2)$$

These functions are depicted in figures 4*a* and 4*b*, respectively. A's epistemic risk function is associated with the ordinary Brier score. Therefore, if A seeks to minimize epistemic risk in the absence of information his credence function will be (0.5, 0.5). Under these conditions, minimizing epistemic risk and applying the ordinary Laplacean POI give the same recommendation. Now consider B. Given her risk function, the risk-free credences are (.42, .58). Given these credences, B would be indifferent between taking a bet on h or its negation. But this credence function is not uniform. In other words, (.42, .58) is the prior credence function recommended by a non-Laplacean indifference principle associated with B's nonsymmetric epistemic risk function.[17]

16. Note that it will not always be possible to infer an agent's risk function and her attitudes to epistemic risk from information about her credences alone. That is, if we learn that the agent's credence for h is, say, 0.75, we may not know whether this is because the agent is an epistemic risk minimizer with a nonsymmetric epistemic risk function or an epistemic risk taker with a symmetric epistemic risk function. A similar screening problem faces someone attempting to read off ordinary risk attitudes from information about preferences. For instance, suppose an agent declines to pay \$1 for a bet that pays \$2 if a certain coin lands on heads and \$0 otherwise. This may be either because the agent is risk averse and assigns equal probabilities to heads and tails or because the agent is risk neutral but believes the coin to be biased toward tails.

17. One might worry that an agent considering her expected epistemic risk could come to the conclusion that she should not adopt risk-free credences because from a perspective of nonuniform credences the risk-free distribution might not minimize risk in expectation. However, we should avoid reference to expected epistemic risk altogether, espe-

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Pettigrew (2016a) argues for the Laplacean POI from considerations of accuracy, as measured by the Brier score, and a minimax decision rule. On the approach we have developed, a more general result follows: the requirement to identify the prior that minimizes epistemic risk under a convex and symmetric risk function will always suggest the same credence function as the ordinary Laplacean POI. But this is not an argument for the uniform prior. Rather, it suggests that we have a family of indifference principles associated with different epistemic risk functions. And whether an agent finds the Laplacean POI attractive depends on her normative judgments regarding the relative severity of approaching different types of graded error.

The notion, due especially to Jaynes (1957), that the right prior is to be found by identifying the maximum entropy distribution is a combination of two separate normative principles: (*a*) that one ought to minimize epistemic risk and (*b*) that one ought to evaluate epistemic risk using a convex, symmetric function. The framework developed here enables us to distinguish the two principles: even if we agree that minimizing epistemic risk is desirable, the appropriate prior may not be uniform. Therefore, the Jaynesian commitment to maximum entropy priors is a commitment to a particular attitude to how much risk is rationally permissible (as little as possible) and how different types of errors are to be evaluated (equally). These are strong normative assumptions that, despite the size of the literature on the problem of the priors and the principle of maximum entropy, had not been adequately addressed.

Appendix

Theorem 1. For strictly concave and twice differentiable entropy function H and risk function R defined on [0, 1], $R(p) + H(p) = k$, where $k = \min_p R(p) = \max_p H(p)$.

Proof. Recall that $R(p) = \int_p^{p^*} |s_1(t) - s_0(t)|\, dt$, where $p^* = \arg\max_{p \in [0,1]} H(p)$. This implies that $H(p^*) = k$, and, given the conditions on entropy, it also implies that $H'(p^*) = 0$ and $H''(p^*) < 0$. These conditions are satisfied if s is strictly proper.

cially in the context of identifying a prior, in part because it is not clear what credences such an agent would use to compute an expectation. Indeed, as noted in sec. 6, the relevant underlying random quantity is accuracy, and we have defined risk as a function of its expectation. Therefore, when we consider risk in identifying a prior, we suppose that the agent is aware of the form of her risk function and ask which credences, if she were to adopt them, would in fact minimize that function.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Existence of risk-free point.—Since $H(p)$ is a strictly concave continuous function that is closed and bounded on [0, 1], the extreme value theorem guarantees that p^* exists.

Duality of risk and entropy.—Savage (1971) shows that we can express $s_v(p)$ in terms of strictly concave and continuous $H(p)$ as follows:

$$s_1(p) = H(p) + (1 - p)H'(p) \qquad s_0(p) = H(p) - pH'(p)$$

Let $h(p) = s_1(p) - s_0(p)$. We can expand $h(p)$ in terms of the entropy $H(p)$:

$$\begin{aligned} h(p) &= [H(p) + (1 - p)H'(p)] - [H(p) - pH'(p)] \\ &= (1 - p)H'(p) + pH'(p) \\ &= H'(p). \end{aligned}$$

Therefore, $\int_p^{p^*} h(t)\,dt = \int_p^{p^*} H'(t)\,dt = H(p^*) - H(p)$. Since $R(p) = \int_p^{p^*} |h(t)|\,dt$, we can use the preceding identity to evaluate $R(p)$ in parts.

For $s_1(p) > s_0(p)$,

$$\begin{aligned} R(p) &= \int_p^{p^*} h(t)\,dt \\ &= H(p^*) - H(p) \\ &= k - H(p) \end{aligned}$$

For $s_0(p) > s_1(p)$,

$$\begin{aligned} R(p) &= -\int_{p^*}^{p} h(t)\,dt \\ &= -[H(p) - H(p^*)] \\ &= k - H(p) \end{aligned}$$

For $s_0(p) = s_1(p)$,

$$\begin{aligned} R(p) &= \int_p^{p^*} h(t)\,dt \\ &= k - k = 0 \end{aligned}$$

QED

Theorem 2. Given a random variable $X : S \to W$, where the underlying scoring rule s is strictly proper, and two cdfs P and Q, if P is a mean-preserving epistemic spread of Q, then $R(P) > R(Q)$.

Proof. Suppose P is a mean-preserving epistemic spread of Q. Then $H(Q) > H(P)$. Let P^* be the risk-free credence function so that $H(P^*) = R(P^*) = 0$. Then given our general expression of epistemic risk in terms of entropic change, $H(P^*) - H(Q) < H(P^*) - H(P)$. Therefore, $R(P) > R(Q)$. QED

Theorem 3. Given a random variable $X : S \to W$, where $W \subseteq \mathbb{R}$ is the image of scoring rule s, let $V = \{P_1, \ldots, P_n\}$ be a set of cdfs for X. Given a random variable $Y : S \to W^*$, where $W^* \subseteq \mathbb{R}$ is the image of scoring rule s^*, let $U = \{Q_1, \ldots, Q_n\}$ be a set of corresponding cdfs for Y. This means that for each outcome $h \in S$, the probability assigned to h by P_i is equal to the probability assigned to h by Q_i, but whereas in the first case the out-

https://doi.org/10.1086/703552 Published online by Cambridge University Press

come h is described by s, in the second case it is described by s^*. Suppose (1) s and s^* are truth-directed scoring rules, whose risk functions R and R^* are such that (2) $R'' > 0$, $R^{*\prime\prime} > 0$, and (3) $\arg\min R = \arg\min R^*$. Then $R(P_i) > R(P_j)$ if and only if $R(Q_i) > R(Q_j)$.

Proof. Sufficiency: assume $R(P_i) > R(P_j)$ for arbitrary $i \neq j$. Recall that $R(P) = E[P^*] - E[P]$, where $P^* = \max_{P \in V} E[P]$ is the risk-free cdf. Conditions 2 and 3, together with the extreme value theorem, imply that P^* exists. Condition 3 implies that $P^* = Q^*$. Finally, condition 1 implies that, if $E[P_i] > E[P_j]$, then $E[Q_i] > E[Q_j]$. Therefore,

$$
\begin{aligned}
& R(P_i) > R(P_j) | \\
& \rightarrow E[P^*] - E[P_i] > E[P^*] - E[P_j] \\
& \rightarrow E[P^* - P_i] > E[P^* - P_j] \\
& \rightarrow E[Q^* - P_i] > E[Q^* - P_j] \\
& \rightarrow E[Q^* - Q_i] > E[Q^* - Q_j] \\
& \rightarrow R(Q_i) > R(Q_j).
\end{aligned}
$$

Necessity: The procedure above is reversible (i.e., the expressions remain true if we swap Q's for P's and W for V). QED

REFERENCES

Arrow, K. J. 1965. *Aspects of the Theory of Risk Bearing*. Helsinki: Yrjõ Jahnsson Sãätiõ.

———. 1971. *Essays in the Theory of Risk Bearing*. Chicago: Markham.

Buchak, L. 2010. "Instrumental Rationality, Epistemic Rationality, and Evidence Gathering." *Philosophical Perspectives* 24 (1): 85–120.

———. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Fallis, D. 2007. "Attitudes toward Epistemic Risk and the Value of Experiments." *Studia Logica* 86 (2): 215–46.

Gibbard, A. 2008. *Rational Credence and the Value of Truth*. Oxford Studies in Epistemology 2. Oxford: Oxford University Press.

Goldman, A. I. 1999. *Knowledge in a Social World*. Oxford: Clarendon.

———. 2002. *Pathways to Knowledge: Private and Public*. Oxford: Oxford University Press.

Greaves, H., and D. Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind* 115 (459): 607–32.

Grunwald, P., and A. Dawid. 2004. "Game Theory, Maximum Entropy, Minimum Discrepancy and Robust Bayesian Decision Theory." *Annals of Statistics* 32 (4): 1367–433.

Horowitz, S. 2018. "Epistemic Utility and the 'Jamesian Goals.'" In *Epistemic Consequentialism*, ed. H. Ahlstrom-Vij and J. Dunn. Oxford: Oxford University Press.

James, W. 1896. "The Will to Believe." *New World* 5 (June): 327–47.

Jaynes, E. T. 1957. "Information Theory and Statistical Mechanics." Pts. 1 and 2. *Physical Review* 106 (4): 620–30; 108 (2): 171–90.

———. 1963/1983. "Brandeis Summer Institute Lectures in Theoretical Physics." In *Papers on Probability, Statistics and Statistical Physics*, ed. R. Rosenkrantz. Dordrecht: Reidel.

———. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

https://doi.org/10.1086/703552 Published online by Cambridge University Press

Jefferson, T. 1785/1832. *Notes on the State of Virginia*. Boston: Lilly & Wait.
Joyce, J. M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65 (4): 575–603.
———. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, ed. F. Huber and C. Shmidt-Petri. Dordrecht: Springer.
———. 2015. "Prior Probabilities as Expressions of Epistemic Values." Unpublished manuscript, University of Michigan, Ann Arbor.
Leitgeb, H., and R. Pettigrew. 2010. "An Objective Justification of Bayesianism." Pt. 1, "Measuring Inaccuracy," and pt. 2, "The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77 (2): 201–35, 236–72.
Levi, I. 1962. "On the Seriousness of Mistakes." *Philosophy of Science* 29 (1): 47–65.
———. 1974. *Gambling with Truth*. Cambridge, MA: MIT Press.
Levinstein, B. 2017. "Permissive Rationality and Sensitivity." *Philosophy and Phenomenological Research* 94 (2): 342–70.
Maher, P. 1990. "Why Scientists Gather Evidence." *British Journal for the Philosophy of Science* 41 (1): 103–19.
———. 1993. *Betting on Theories*. Cambridge: Cambridge University Press.
Peirce, C. S. 1879/1967. "Note on the Theory of the Economy of Research." Technical report, US Coast Survey, US Government Publishing Office. Repr. in *Operations Research* 15 (4): 643–48.
Pettigrew, R. 2016a. "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research* 92 (1): 35–59.
———. 2016b. "Jamesian Epistemology Formalised: An Explication of 'The Will to Believe.'" *Episteme* 13 (3): 253–68.
Pratt, J. W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32 (1/2): 122–36.
Pritchard, D. 2007. "Anti-luck Epistemology." *Synthese* 158 (3): 277–97.
———. 2017. "Epistemic Risk." *Journal of Philosophy* 113 (11): 550–71.
Rescher, N. 1976. "Peirce and the Economy of Research." *Philosophy of Science* 43 (1): 71–98.
Rothschild, M., and J. E. Stiglitz. 1970. "Increasing Risk." Pt. 1, "A Definition." *Journal of Economic Theory* 2 (3): 225–43.
Savage, L. J. 1954. *The Foundations of Statistics*. New York: Dover.
———. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (336): 783–801.
Seidenfeld, T. 1986. "Entropy and Uncertainty." *Philosophy of Science* 53 (4): 467–91.
Selten, R. 1998. "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics* 1 (1): 43–61.
Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell Systems Technical Journal* 27 (3): 379–423; 27 (4): 623–66.
van Fraassen, B. C. 1984. "Belief and the Will." *Journal of Philosophy* 81 (5): 235–56.
———. 1989. *Laws and Symmetry*. Oxford: Clarendon.
von Neumann, J., and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
Williamson, J. 2010. *In Defense of Objective Bayesianism*. Oxford: Oxford University Press.

https://doi.org/10.1086/703552 Published online by Cambridge University Press