Fundamentals of Bayesian Epistemology 2

Arguments, Challenges, Alternatives

MICHAEL G. TITELBAUM

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2022

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2021949533

ISBN 978–0–19–870760–8 (hbk.)
ISBN 978–0–19–870761–5 (pbk.)

DOI: 10.1093/oso/9780198707608.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

VOLUME 2

III. APPLICATIONS

IV. ARGUMENTS FOR BAYESIANISM

10
Accuracy Arguments

The previous two chapters considered arguments for probabilism based on Representation Theorems and Dutch Books. We criticized both types of argument for beginning with premises about practical rationality—premises about how a rational agent views certain acts (especially acts of placing bets). We want to establish the probability axioms as requirements of *theoretical* rationality on an agent's credences, and it's difficult to move from practical premises to a theoretical conclusion.

This chapter builds arguments for probabilism from explicitly epistemic premises. The basic idea is that, as representational attitudes, credences can be assessed for **accuracy**. We often assess other doxastic attitudes, such as binary beliefs, for accuracy: a belief in the proposition P is accurate if P is true; disbelief in P is accurate if P is false. A traditional argument moves from such accuracy assessments to a rational requirement that agents' belief sets be logically consistent (Chapter 1's Belief Consistency norm). The argument begins by noting that if a set of propositions is logically inconsistent, then by definition there is no (logically) possible world in which all of those propositions are true. So if an agent's beliefs are logically inconsistent, she's in a position to know that at least some of them are inaccurate. Moreover, she can know this a priori—without invoking any contingent truths. (Since an inconsistent set contains falsehoods in *every* possible world, no matter which world is actual her inconsistent belief set misrepresents how things are.) Such unavoidable, a priori inaccuracy reveals a rational flaw in any logically inconsistent set of beliefs.[1]

There are plenty of potential flaws in this argument—starting with its assumption that beliefs have a teleological "aim" of being accurate. But the argument is a good template for the arguments for probabilism to be discussed in this chapter. Whatever concerns you have about the Belief Consistency argument above, keep them in mind as you consider accuracy arguments for probabilism.

Assessing credences for accuracy isn't as straightforward as assessing binary beliefs: a credence of, say, 0.6 in proposition P doesn't say that P is true, but neither does it say that P is false. So we can't assess the accuracy of this credence

Fundamentals of Bayesian Epistemology 2: Arguments, Challenges, Alternatives. Michael G. Titelbaum, Oxford University Press. © Michael G. Titelbaum 2022. DOI: 10.1093/oso/9780192863140.003.0010

by asking whether it assigns a truth-value to P matching P's truth-value in the world. Nor can we say that $\text{cr}(P) = 0.6$ is accurate just in case P is true "to degree 0.6"; we've assumed that propositions are wholly true or wholly false, full stop. So just as we moved from classificatory to quantitative doxastic attitudes in Chapter 1, we need to move from a classificatory to a quantitative concept of accuracy. This chapter will begin by considering various numerical measures of just *how* accurate a credence (or set of credences) is. We'll start with historical "calibration" approaches that measure credal accuracy by comparing credences with frequencies. Then we'll reject calibration in favor of contemporary "gradational accuracy" approaches.

The most commonly used gradational accuracy measure is known as the Brier score. Using the Brier score, we will construct an argument for probabilism similar to the Belief Consistency argument above: violating the probability axioms impedes a credence set's accuracy in every possible world. It will then turn out that an argument like this can be constructed using not just the Brier score, but any gradational accuracy measure in a class known as the "strictly proper scoring rules".

Which leads to the question of why strictly proper scoring rules are superior to other accuracy-measurement options—especially options that rule *out* probabilism. The spectre will arise once more that our argument for probabilism has a question-begging, Linearity-In, Linearity-Out structure. This will force us to ask something you may have started wondering over the last couple of chapters: How important is it, *really*, that rational credences satisfy Finite Additivity, as opposed to other norms with similar consequences for thought and behavior?

Besides arguing for probabilism, Bayesian epistemologists have offered accuracy-based arguments for other norms such as the Principal Principle (Pettigrew 2013a), the Principle of Indifference (Pettigrew 2014), Reflection (Easwaran 2013), and Conglomerability (Easwaran 2013). We'll close this chapter with an argument for Conditionalization based on minimizing expected future inaccuracy. Unfortunately this argument has the same drawback as Dutch Strategy arguments for Conditionalization: it is insufficient on its own to establish truly *diachronic* norms.

10.1 Accuracy as calibration

In Section 5.2.1 we briefly considered a putative rational principle for matching one's credence that a particular outcome will occur to the frequency with which that kind of outcome occurs. In that context, the match was supposed to be between one's credence that outcome B will occur and the frequency with which *one's evidence* suggests B-type outcomes occur. But we might instead assess an agent's credences relative to *actual frequencies* in the world: If events of type A actually produce outcomes of type B with frequency x, an agent's credence that a particular A-event will produce a B-outcome is more accurate the closer it is to x.

Now imagine that an agent managed to be perfectly accurate with respect to the actual frequencies. In that case, she would assign credence 2/3 to types of outcomes that occurred 2/3 of the time, credence 1/2 to outcome-types that occurred 1/2 of the time, etc. Or—flipping this around—propositions to which she assigned credence 2/3 would turn out to be true 2/3 of the time, propositions to which she assigned credence 1/2 would turn out to be true 1/2 of the time, etc. This conception of accuracy—getting the frequencies right, as it were—motivates assessing credences with respect to their

Calibration: A credence distribution over a finite set of propositions is perfectly calibrated when, for any real x, out of all the propositions to which the distribution assigns x, the fraction that turn out to be true is x.

For example, suppose your weather forecaster comes on television every night and reports her degree of confidence that it will snow the next day. You might notice that every time she says she's 20% confident of snow, it snows the next day. In that case she's not a very accurate forecaster. But if it snows on just about 20% of those days, we'd say she's doing her job well. If exactly 20% of the days on which she's 20% confident of snow turn out to have snow (and exactly 30% of the days on which she's 30% confident... etc.), we say the forecaster is perfectly calibrated.[2] Calibration seems a plausible way to gauge accuracy.[3]

I've defined only what it means to be *perfectly* calibrated; there are also numerical measures for assessing degrees of calibration short of perfection (see Murphy 1973).[4] But all the good and bad consequences of reading accuracy as calibration can be understood by thinking solely about perfect calibration. First, the good: van Fraassen (1983) and Abner Shimony (1988) both argued for probabilism by showing that in order for a credence distribution to be embeddable in larger and larger distributions with calibration scores approaching perfection, that original credence distribution must satisfy the probability axioms. This seems a powerful argument for probabilism—if we're on board with calibration as a measure of accuracy.

Here's why we might not be. Consider two agents, Sam and Diane, who assign the following credence distributions over propositions X_1 through X_4:

	X_1	X_2	X_3	X_4
Sam	1/2	1/2	1/2	1/2
Diane	1	1	1/10	0

Now suppose that propositions X_1 and X_2 are true, while X_3 and X_4 are false. Look at the table and ask yourself whose credences intuitively seem more accurate.[5]

I take it the answer is Diane. Yet Sam's credences are perfectly calibrated—he assigns credence 1/2 to all four propositions, exactly half of which are true—while Diane's credences are not. This is an intuitive flaw with measuring accuracy by calibration.

A similar point can be made by considering the following (real life!) example: On the morning of February 1, 2015, I looked outside and found it was snowing heavily. At least four inches had accumulated during the night, the snow was still coming down, and it showed no signs of stopping. The online weather report on my smartphone, however, showed an at-the-moment 90% probability of snow. Why hadn't the forecaster simply looked out her window and updated the report to 100%?

I was suddenly struck by a possible explanation. Let's imagine (what's probably not true) that the forecaster posts to the online weather report her current credence that it will snow on the current day. Suppose also that weather forecasting sites are graded for accuracy, and promoted on search engines based on how well they score. Finally, suppose this accuracy scoring is done by measuring calibration. What if, up to February 1, it had snowed every time the forecaster reported a 100% credence, but it had snowed on only eight of the nine occasions on which she had expressed a 90% credence? The snow on February 1 would then present her with an opportunity. She could report her true, 100% confidence in snow for February 1 on the website. Or she could post a 90% probability of snow. Given that it was clearly snowing on February 1, the latter option would bring her up to a perfect calibration score, and shoot her website to the top of the search rankings. Calibration gives the forecaster an incentive to misreport her own credences—and the content of her own evidence.

Calibration is one example of a **scoring rule**; a procedure for rating distributions with respect to accuracy. James M. Joyce reports that "the term 'scoring rule' comes from economics, where values of [such rules] are seen as imposing penalties for making inaccurate probabilistic predictions" (2009, p. 266). Done right, the imposition of such penalties can be a good way of finding out what experts really think—what's known as **credence elicitation**. If you reward (or punish) an expert according to the accuracy of her reports, you incentivize her to gather the best evidence she can, consider it carefully, and then report to you her genuine conclusions. Seen through this lense of credence elicitation, calibration fails as a scoring rule. As we've just seen, rewarding a forecaster according to her level of calibration can incentivize her to misreport her true opinions, and what she takes to be the import of her evidence.

Yet perhaps it's unfair to criticize calibration on the grounds that it perversely incentivizes credence *reports*; norms for assertion can be messy, and anyway probabilism is a norm on agents' thoughts, not their words. So let's consider calibration as a direct accuracy measure of our forecaster's credences. Prior to February 1 it has snowed whenever the forecaster was certain of snow, but on the days when she assigned 0.9 credence to snow, it has snowed eight of nine times. Looking out her window and seeing snow, the forecaster assigns credence 1 to snow.[6] Yet if her goal is to be as accurate as possible with her credences, and if accuracy is truly measured by calibration, then the forecaster will *wish* that her credence in snow was 0.9. After all, that would make her pefectly calibrated!

Assessing the forecaster's credences by calibration makes those credences *unstable*. By the forecaster's own lights—given the credences she has formed in light of her evidence—she thinks she'd be better off with different credences. Such instability is an undesirable feature in a credence distribution, and is generally thought to be a hallmark of irrationality. David Lewis offers the following analogy:

> It is as if *Consumer Bulletin* were to advise you that *Consumer Reports* was a best buy whereas *Consumer Bulletin* itself was not acceptable; you could not possibly trust *Consumer Bulletin* completely thereafter. (1971, p. 56)

If we use calibration to measure accuracy, the weather forecaster's credence distribution becomes unstable. Such instability is a sign of irrationality. So from a calibration point of view, there's something rationally wrong with the forecaster's credences. But in reality there's nothing wrong with the forecaster's credences—they are a perfectly rational response to the evidence before her eyes! The problem lies with calibration as a measure of accuracy; calibration renders unstable credence distributions that are rationally permissible (if not rationally required!).

There are further ways in which calibration rewards agents for ignoring their evidence. Notice that any agent assigning credences over a partition of n propositions can secure a perfect calibration score by assigning each proposition a credence of $1/n$. For instance, if a six-sided die is to be rolled, an agent can guarantee herself perfect calibration (no matter how the roll comes out!) by assigning each possible outcome a credence of 1/6. Depending on how you feel about the Principle of Indifference (Section 5.3), this might be a reasonable assignment when the agent has no evidence relevant to the members of the partition. But now suppose the agent gains highly reliable evidence that the die is biased in favor of coming up 6. Altering her credences to reflect that bias won't earn her a better calibration score than the uniform 1/6 distribution, and might very well serve her worse.

One could make various attempts here to save calibration as a plausible measure of accuracy. For instance, calibration scores are less easily manipulable if we measure them only in the long-run. But this generates questions about the accuracy of credences in non-repeatable events, and soon we're assessing not actual long-run calibration but hypothetical calibration in the limit. Before long, we've made all the desperate moves used to prop up the frequency theory of "probability" (Section 5.1.1), and run into all the same problems.

The response here should be the same as it was with the frequency theory: Rather than deploy a notion that emerges only when events are situated in a larger collective, we find a notion (like propensity) that can be meaningfully applied to single cases considered one at a time. Looking back at Sam and Diane, our intuitive judgment that Diane is globally more accurate than Sam arises from local judgments that she was more accurate than him on each individual proposition. If you knew only the truth-value of X_1, you could still have said that Diane was more accurate than Sam on that one proposition. Our accuracy intuitions apply piece-wise; we assess credences one proposition at a time, then combine the results into a global accuracy measure.

10.2 The gradational accuracy argument for probabilism

10.2.1 The Brier score

We will now develop what's known as the "gradational accuracy" approach to evaluating credences. Our guiding idea will be that inaccuracy is distance from truth—a credence distribution gains accuracy by moving its values closer to the truth-values of propositions. Of course, credence values are real numbers,

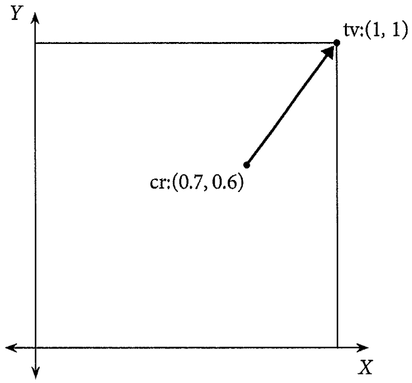

Figure 10.1 The Brier score

while truth-values are not. But it's natural to overcome that obstacle by letting 1 stand for truth and 0 stand for falsehood. Just as we have a distribution cr expressing the agent's credences in propositions, we'll have another distribution tv reflecting the truth-values of those propositions. Distribution tv assigns numerical values to the propositions in $\mathcal{L}$ such that $\text{tv}(X) = 1$ if X is true and $\text{tv}(X) = 0$ if X is false.[7]

Once we have distribution cr representing the agent's credences and distribution tv representing the truth, we want a scoring rule that measures how far apart these distributions are from each other. It's easiest to visualize the challenge on a diagram. To simplify matters, consider a credence distribution over only two propositions, X and Y. Our agent assigns $\text{cr}(X) = 0.7$ and $\text{cr}(Y) = 0.6$. I have depicted this credence assignment in Figure 10.1. In this diagram the horizontal axis represents the proposition X while the vertical axis represents Y. Any credence assignment to these two propositions can be represented as an ordered pair; I have placed a dot at the agent's cr-distribution of $(0.7, 0.6)$.

What about the values of tv? Let's suppose that propositions X and Y are both true. So $\text{tv}(X) = \text{tv}(Y) = 1$. I have marked $(1, 1)$—the location of tv on the diagram—with another dot. Now our question is how to measure the inaccuracy of the agent's credences; how should we gauge how far cr is from tv?

A natural suggestion is to use distance as the crow flies, indicated by the arrow in Figure 10.1. A quick calculation tells us that the length of the arrow is:

$$(1 - 0.7)^2 + (1 - 0.6)^2 = (0.3)^2 + (0.4)^2 = 0.25 \qquad (10.1)$$

Pythagorean Theorem aficionados will note the lack of a square-root in this distance expression (the arrow is actually 0.5 units long). For the time being, we're going to use inaccuracy measurements only for ordinal comparisons (which credence distribution is *farther* from the truth), so particular numerical values don't matter much—and neither does the square-root.

When generalized to a credence distribution over finitely many propositions $\{X_1, X_2, \ldots, X_n\}$, this distance measure of inaccuracy becomes

$$\begin{aligned} I_{BR}(\mathrm{cr}, \omega) = (\mathrm{tv}_\omega(X_1) - \mathrm{cr}(X_1))^2 + (\mathrm{tv}_\omega(X_2) - \mathrm{cr}(X_2))^2 + \ldots \\ + (\mathrm{tv}_\omega(X_n) - \mathrm{cr}(X_n))^2 \end{aligned} \tag{10.2}$$

A few notes about this equation: First, what are the ωs doing in there? We usually want to evaluate the inaccuracy of your credence distribution relative to conditions in the actual world. But sometimes we'll wonder how inaccurate your credences would've been if you'd maintained your distribution but lived in a different possible world. For example, in Figure 10.1 we might wonder how inaccurate the credence distribution cr would have been had X and Y both been false. That is, we might want to calculate the distance between cr and the point $(0, 0)$. Equation (10.2) calculates the inaccuracy of credence distribution cr in an *arbitrary* possible world ω. $\mathrm{tv}_\omega(X_i)$ represents the truth-value of proposition X_i in world ω; $I_{BR}(\mathrm{cr}, \omega)$ then measures the inaccuracy of cr relative to conditions in that world. (So for the credence distribution $(0.7, 0.6)$ and the world $(0, 0)$, Equation (10.2) would yield an I_{BR}-value of $0.7^2 + 0.6^2 = 0.85$.)[8]

Second, Equation (10.2) tallies up inaccuracy one proposition at a time, then sums the results. For any credence distribution cr and particular proposition X_i, evaluating $(\mathrm{tv}(X_i) - \mathrm{cr}(X_i))^2$ is a way of gauging how far off distribution cr is *on that particular proposition*. Equation (10.2) makes that calculation for each individual proposition X_i, then adds up the results. In general, a scoring rule that sums the results of separate calculations made on individual propositions is called **separable**. Separable scoring rules track our intuition that accuracy assessments of an entire credence distribution can be built up piece-wise, considering the accuracy of one credence at a time; this was exactly the feature we found lacking in calibration's evaluation of Sam and Diane.[9]

The scoring rule described by Equation (10.2) is known as the Euclidean distance, the quadratic loss function, or most commonly the **Brier score.**[10] (This accounts for the "BR" subscript in I_{BR}.) The Brier score is hardly the only scoring rule available, but it is natural and widely used. So we will stick with it for the time being, until we examine other options in Section 10.3. At that point we'll find that even among the separable scoring rules, there may be ordinal non-equivalence—two separable scoring rules may disagree about which distribution is more accurate in a given world. Nevertheless, all the separable scoring rules have some features in common. For instance, while $I_{BR}(\text{cr}, \omega)$ is in some sense a global measure of the inaccuracy of cr in world ω, it doesn't take into account any wholistic or interactive features among the individual credences cr assigns. Separable scores can't, for example, take into account the sum or difference of $\text{cr}(X_i)$ and $\text{cr}(X_j)$ for $i \neq j$.[11]

Finally, I_{BR} and the other gradational measures we'll consider calculate the *in*accuracy of credence distributions in particular worlds. So an agent striving to be as accurate as possible will seek to *minimize* her score. Some authors prefer to work with credence distributions' **epistemic utility**, a numerical measure of epistemic value that rational agents *maximize*. Now there may be many aspects of a credence distribution that make it epistemically valuable or disvaluable besides its distance from the truth. But many authors work under the assumption that accuracy is the sole determiner of a distribution's epistemic value, in which case that value can be calculated directly from the distribution's inaccuracy. (The simplest way is to let the epistemic utility of distribution cr in world ω equal $1 - I_{BR}(\text{cr}, \omega)$.) If you find yourself reading elsewhere about accuracy arguments, be sure to notice whether the author asks agents to *minimize inaccuracy* or *maximize utility*. On either approach, the best credence is the one closest to the pin (the distribution tv). But with inaccuracy, as in golf, lowest score wins.

10.2.2 Joyce's accuracy argument for probabilism

In our discussion of calibration we said that it's rationally problematic for an agent's credence distribution to be "unstable"—for it to seem to the agent, by her own lights, that another credence distribution would be preferable to her own. We ultimately rejected assessing agents' credences using calibration, but now we have an alternative accuracy measure: the Brier score. If we could convince an agent that her credences are less accurate, as measured by the Brier score, than some other distribution over the same set of propositions, then it would seem irrational for her to maintain her credence distribution (as opposed to the other one).

How can we convince an agent that her credences are less accurate than some alternative? Inaccuracy is always measured *relative to a world*. Presumably the agent is interested how things stand in the *actual* world, but presumably she also has some uncertainty as to which propositions are true or false in the actual world. If she doesn't know the tv-values in the actual world, she won't be able to calculate her own Brier score in that world, much less the score of an alternative distribution.

But what if we could show her that there exists a single distribution that is more accurate than her own *in every logically possible world*? Then she wouldn't need to know which possible world was actual; she could determine on an a priori basis that however things stand in the actual world, she would be more accurate if she had that other distribution. In light of information like this, her present credences would look irrational. This line of thought is enshrined in the following principle:

Admissibles Not Dominated: If an agent's credence distribution is rationally permissible, and accuracy is measured with an acceptable scoring rule, then there does not exist another distribution that is more accurate than hers in every possible world.

Admissibles Not Dominated is a conditional. Contraposing it, we get that any credence distribution accuracy-dominated by another distribution on an acceptable scoring rule is rationally impermissible (or "inadmissible", in the accuracy literature's jargon).

Repurposing a theorem of de Finetti's (1974), and following on the work of Rosenkrantz (1981), Joyce (1998) demonstrated the

Gradational Accuracy Theorem: Given a credence distribution cr over a finite set of propositions $\{X_1, X_2, \ldots, X_n\}$, if we use the Brier score $I_{BR}(\mathrm{cr}, \omega)$ to measure inaccuracy then:

- If cr does *not* satisfy the probability axioms, then there exists a probabilistic distribution cr′ over the same propositions such that $I_{BR}(\mathrm{cr}', \omega) < I_{BR}(\mathrm{cr}, \omega)$ in every logically possible world ω; and
- If cr *does* satisfy the probability axioms, then there does not exist any cr′ over those propositions such that $I_{BR}(\mathrm{cr}', \omega) < I_{BR}(\mathrm{cr}, \omega)$ in every logically possible world.

The Gradational Accuracy Theorem has two parts. The first part says that if an agent has a non-probabilistic credence distribution cr, we will be able to find a probabilistic distribution cr′ defined over the same propositions as cr that

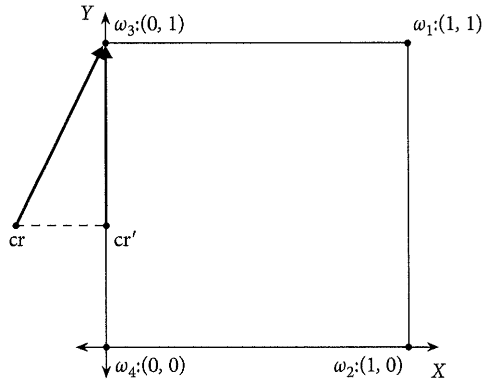

Figure 10.2 Violating Non-Negativity

accuracy-dominates cr. No matter what the world is like, distribution cr′ will be less inaccurate than cr. So the agent with distribution cr can be certain that, come what may, she is leaving a certain amount of accuracy on the table by assigning cr rather than cr′. There's a cost in accuracy, independent of what you think the world is like and therefore discernible a priori, to assigning a non-probabilistic credence distribution—much as there's a guaranteed accuracy cost to assigning logically inconsistent beliefs. On the other hand (and this is the second part of the theorem), if an agent's credence distribution *is* probabilistic, then no distribution (probabilistic or otherwise) is more accurate in every possible world. This seems a strong advantage of probabilistic credence distributions.[12]

Proving the second part of the theorem is difficult, but I will show how to prove the first part. There are three probability axioms—Non-Negativity, Normality, and Finite Additivity—so we need to show how violating each one leaves a distribution susceptible to accuracy domination. We'll take them one at a time, in order.

Suppose credence distribution cr violates Non-Negativity by assigning some proposition a negative credence. In Figure 10.2 I've imagined that cr assigns credences to two propositions, X and Y, bearing no special logical relations to each other. cr violates Non-Negativity by assigning $\text{cr}(X) < 0$. (The value of $\text{cr}(Y)$ is irrelevant to the argument, but I've supposed it lies between 0 and 1.) We introduce probabilistic cr′ such that $\text{cr}'(Y) = \text{cr}(Y)$ but $\text{cr}'(X) = 0$; cr′ is the closest point on the Y-axis to distribution cr.

We need to show that cr′ is less inaccurate than cr no matter which possible world is actual. Given our two propositions X and Y, there are four possible worlds.[13] I've marked them on the diagram as ω_1, ω_2, ω_3, and ω_4, determining

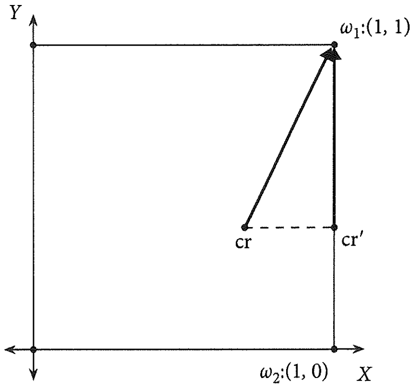

Figure 10.3 Violating Normality

the coordinates of each world by the truth-values it assigns to X and Y. (In ω_2, for instance, X is true and Y is false.) We now need to show that for each of these worlds, cr′ receives a lower Brier score than cr. In other words, we need to show that cr′ is closer to each world as the crow flies than cr is.

Clearly cr′ is closer to ω_2 and ω_1 than cr is, so cr′ is less inaccurate than cr relative to both ω_2 and ω_1. What about ω_3? I've indicated the distances from cr and cr′ to ω_3 with arrows. Because cr′ is the closest point on the Y-axis to cr, the points cr, cr′, and ω_3 form a right triangle. The arrow from cr to ω_3 is the hypotenuse of that triangle, while the arrow from cr′ to ω_3 is a leg. So the latter must be shorter, and cr′ is less inaccurate by the Brier score relative to ω_3. A parallel argument shows that cr′ is less inaccurate relative to ω_4. So cr′ is less inaccurate than cr relative to each possible world.

That takes care of Non-Negativity.[14] The accuracy argument against violating Normality is depicted in Figure 10.3. Suppose X is a tautology and cr assigns it some value other than 1. Since X is a tautology, there are no logically possible worlds in which it is false, so there are only the worlds ω_2 and ω_1 to consider. We construct cr′ such that $\text{cr}'(Y) = \text{cr}(Y)$ and $\text{cr}'(X) = 1$. cr′ is closer than cr to ω_1 because the arrow from cr to ω_1 is the hypotenuse of a right triangle of which the arrow from cr′ to ω_1 is one leg. A similar argument shows that cr′ is closer than cr to ω_2, demonstrating that cr′ is less inaccurate than cr in every logically possible world.

Explaining how to accuracy-dominate a Finite Additivity violator requires a three-dimensional argument sufficiently complex that I will leave it for an endnote.[15] But we can show in two dimensions what happens if you violate one of the rules that follows from Finite Additivity, namely our Negation rule.

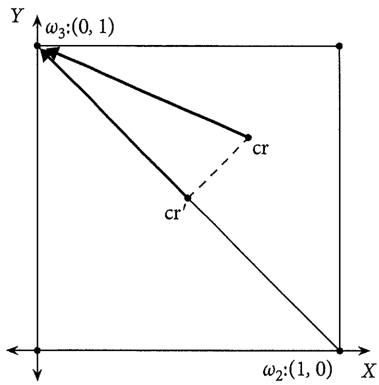

Figure 10.4 Violating Negation

Suppose your credence distribution assigns cr-values to two propositions X and Y such that X is the negation of Y. If you violate Negation, you'll have $\text{cr}(Y) \neq 1 - \text{cr}(X)$.

I've depicted only ω_2 and ω_3 in Figure 10.4 because only those two worlds are logically possible in this case (since X and Y must have opposite truth-values). The diagonal line connecting ω_2 and ω_3 has the equation $Y = 1 - X$; it contains all the credence distributions satisfying Negation. If cr violates Negation, it will fall somewhere off of this line. Then we can accuracy-dominate cr with the point closest to cr lying on the diagonal (call that point cr′). Once more, we've created a right triangle with cr, cr′, and one of our possible worlds. The arrow representing the distance from cr to ω_3 is the hypotenuse of this triangle, while the arrow from cr′ to ω_3 is its leg. So cr′ has the shorter distance, and cr′ is less inaccurate in ω_3 than cr according to the Brier score. A parallel argument applies to ω_2, so cr′ is less inaccurate than cr in each of the two logically possible worlds.[16]

Joyce (1998, 2009) leverages the advantage of probabilistic credence distributions displayed by the Gradational Accuracy Theorem into an argument for probabilism:

<u>Gradational Accuracy Argument for Probabilism</u>

(Premise 1) A rationally permissible credence distribution cannot be accuracy-dominated on any acceptable scoring rule.

(Premise 2) The Brier score is an acceptable scoring rule.

(Theorem) If we use the Brier score, then any non-probabilistic credence distribution can be accuracy-dominated.

(Conclusion) All rationally permissible credence distributions satisfy the probability axioms.

The first premise of this argument is Admissibles Not Dominated. The theorem is the Gradational Accuracy Theorem. The conclusion of this argument is probabilism.

10.3 Objections to the accuracy argument for probabilism

Unlike Representation Theorem and Dutch Book Arguments, the Gradational Accuracy Argument for Probabilism has nothing to do with an agent's decision-theoretic preferences over practical acts. It clearly pertains to the *theoretical* rationality of credences assigned in pursuit of an *epistemic* goal: accuracy. (This is why Joyce's (1998) paper was titled "A Nonpragmatic Vindication of Probabilism".) This is a major advantage of the accuracy argument for probabilism. Of course, one has to be comfortable with the idea that belief-formation is a goal-directed activity—teleological, so to speak—and commentators have objected to that idea. (Examples appear in the Further Reading.)

But I want to focus on a more technical objection that has been with the gradational accuracy approach from its inception. Premise 2 of the Gradational Accuracy Argument states that the Brier score is an acceptable scoring rule. The Brier score is certainly not the only scoring rule possible; why do we think it's acceptable? And what does it even *mean* for a scoring rule to be acceptable in this context?

10.3.1 The absolute-value score

In his original (1998) presentation of the accuracy argument, Joyce selected the Brier score because it exhibits a number of appealing formal properties—what we might think of as adequacy conditions for an acceptable scoring rule. We've already seen that the Brier score is a separable rule. The Brier score also displays

Truth-Directedness: If a distribution cr is altered by moving at least one $\mathrm{cr}(X_i)$ value closer to $\mathrm{tv}_\omega(X_i)$, and no individual cr-values are moved farther away from tv_ω, then $\mathrm{I}(\mathrm{cr}, \omega)$ decreases.

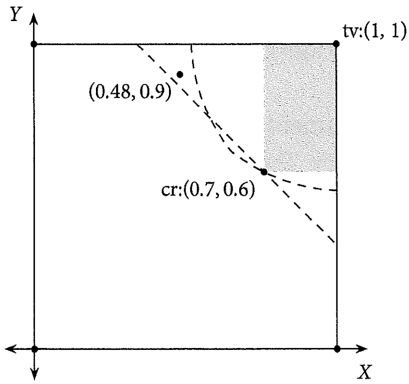

Figure 10.5 Truth-Directedness

The intuitive idea of Truth-Directedness is that if you change your credence distribution by moving some propositions closer to their truth-values, and leaving the rest alone, this should decrease inaccuracy. This condition is depicted in Figure 10.5. (Ignore the dashed elements in that diagram for now.) Assume once more that the agent assigns credences only to the propositions X and Y, and that both these propositions are true in the actual world. If the agent's credence distribution is $(0.7, 0.6)$, every point on or in the gray box (except for $(0.7, 0.6)$ itself) assigns an X-credence or a Y-credence closer to 1 than hers. On a truth-directed scoring rule, all of those distributions are more accurate than the agent's.

The Brier score isn't the only truth-directed scoring rule, or the only way of measuring distance on a diagram. Brier measures distance as the crow flies. But suppose you had to travel from the distribution $(0.7, 0.6)$ to the truth $(1, 1)$ by traversing a rectangular street grid, which permitted movement only parallel to the axes. The shortest distance between those two points measured in this fashion—what's sometimes called the "taxicab distance"—is

$$|1 - 0.7| + |1 - 0.6| = 0.3 + 0.4 = 0.7 \tag{10.3}$$

I've illustrated this distance in Figure 10.6.

Generalizing the taxicab calculation to a distribution over finitely many propositions $\{X_1, X_2, \ldots, X_n\}$ yields

$$\mathrm{I}_{\text{ABS}}(\mathrm{cr}, \omega) = |\mathrm{tv}_\omega(X_1) - \mathrm{cr}(X_1)| + |\mathrm{tv}_\omega(X_2) - \mathrm{cr}(X_2)| + \ldots + |\mathrm{tv}_\omega(X_n) - \mathrm{cr}(X_n)| \tag{10.4}$$

We'll call this the absolute-value scoring rule.

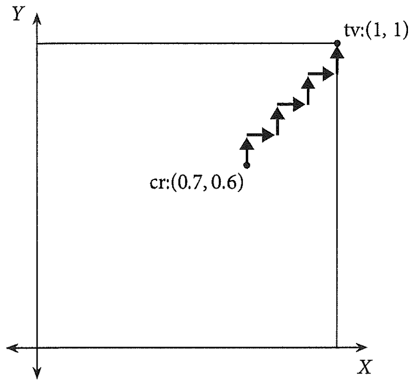

Figure 10.6 The absolute-value score

Both the absolute-value score and the Brier score satisfy Truth-Directedness. We can see this by attending to the dashed elements in Figure 10.5. The dashed line passing through (0.7, 0.6) shows distributions that have the exact *same* inaccuracy as (0.7, 0.6) if we measure inaccuracy by the absolute-value score.[17] Any point between that dashed line and (1, 1) is *more* accurate than (0.7, 0.6) by the absolute-value score. Notice that all the points in the gray box fall into that category, so the absolute-value score is truth-directed.

The dashed quarter-circle shows distributions that are just as inaccurate as (0.7, 0.6) if we measure inaccuracy by the Brier score. Points between the dashed quarter-circle and (1, 1) are less inaccurate than (0.7, 0.6) according to the Brier score. Again, the gray box falls into that region, so the Brier score is truth-directed.

We can see in Figure 10.5 that the Brier score and the absolute-value score are ordinally non-equivalent measures of inaccuracy. To bring out the contrast, consider the distribution (0.48, 0.9). Notice that Truth-Directedness doesn't settle whether this distribution is more or less accurate than (0.7, 0.6)—given that both *X* and *Y* have truth-values of 1, (0.48, 0.9) does better than (0.7, 0.6) with respect to *Y* but worse with respect to *X*. We have to decide whether the *Y* improvement is dramatic enough to merit the *X* sacrifice; Truth-Directedness offers no guidance concerning such tradeoffs. The Brier score and absolute-value score render opposite verdicts on this point. (0.48, 0.9) lies inside the dashed line, so the absolute-value score evaluates this distribution as *less* inaccurate than (0.7, 0.6). But (0.48, 0.9) lies outside the quarter-circle, so the Brier score evaluates it as *more* inaccurate. Here we have a concrete case in which the absolute and Brier scores disagree in their accuracy rankings of two distributions.

Such disagreement is especially important when it comes to the Gradational Accuracy Argument. A Gradational Accuracy Theorem cannot be proven for the absolute-value score; in fact, replacing the Brier score with the absolute-value score in the statement of that theorem yields a falsehood. (We'll demonstrate this in the next section.) This makes the second premise of the Gradational Accuracy Argument crucial. The first premise says that rational credence distributions are not dominated on any acceptable scoring rule. If all the acceptable scoring rules were like the absolute-value score, then nonprobabilistic distributions would not be dominated and the argument could not go through. But if we can establish that the Brier score is acceptable, then we have an argument for probabilism.

10.3.2 Proper scoring rules

How do we decide whether the Brier score or the absolute-value score (or both, or neither) is an acceptable measure of inaccuracy? In his (1998), Joyce offered adequacy conditions beyond Truth-Directedness and separability that favored the Brier score over the absolute-value score. Maher (2002), however, argued that these properties were implausible as requirements on rationally acceptable scoring rules, and defended the absolute-value score. So we're left wondering how to select one over the other.

Historically, the Brier score was favored over the absolute-value score because Brier belongs to a broad class of scoring rules called the "proper" scoring rules. To understand this notion of propriety, we first need to understand *expected* inaccuracies.

Suppose I want to assess the inaccuracy of my friend Rita's credence distribution. We'll simplify matters by stipulating that Rita assigns only two credence values, $\text{cr}_R(X) = 0.7$ and $\text{cr}_R(Y) = 0.6$. Stipulate also that I am going to use the absolute-value score for inaccuracy measurement. We know from Equation (10.3) that if X and Y are both true, Rita's I_{ABS} score is 0.7. The trouble is, I'm not certain whether X or Y is true; I assign positive credence to each of the four truth-value assignments over X and Y. The table below shows my credence distribution (cr) over the four possibilities—which is distinct from Rita's:

	X	Y	cr	$\text{I}_{\text{ABS}}(\text{cr}_R, \cdot)$
ω_1	T	T	0.1	0.7
ω_2	T	F	0.2	0.9
ω_3	F	T	0.3	1.1
ω_4	F	F	0.4	1.6

The last column in this table shows the inaccuracy of Rita's distribution in each of the four possible worlds according to the absolute-value score. If X and Y are both true, her inaccuracy is 0.7; if X is true but Y is false, it's 0.9; etc.

The table tells me the inaccuracy of Rita's distribution in each possible world. I can't calculate her actual inaccuracy, because I'm not certain which possible world is actual. But I can calculate how inaccurate I *expect* Rita's distribution to be. The inaccuracy of a credence distribution is a numerical quantity, and just like any numerical quantity I may calculate my expectation for its value. My expectation for the $\mathrm{I_{ABS}}$ value of Rita's distribution cr_R is:

$$\begin{aligned}\mathrm{EI_{cr}}(\mathrm{cr}_R) &= \mathrm{I_{ABS}}(\mathrm{cr}_R, \omega_1) \cdot \mathrm{cr}(\omega_1) + \mathrm{I_{ABS}}(\mathrm{cr}_R, \omega_2) \cdot \mathrm{cr}(\omega_2) \\ &\quad + \mathrm{I_{ABS}}(\mathrm{cr}_R, \omega_3) \cdot \mathrm{cr}(\omega_3) + \mathrm{I_{ABS}}(\mathrm{cr}_R, \omega_4) \cdot \mathrm{cr}(\omega_4) \\ &= 0.7 \cdot 0.1 + 0.9 \cdot 0.2 + 1.1 \cdot 0.3 + 1.6 \cdot 0.4 = 1.22\end{aligned} \tag{10.5}$$

For each world, I calculate how inaccurate cr_R would be in that world, then multiply by my credence cr that that world is actual.[18] Finally, I sum the results across all four worlds. Notice that because I'm more confident in, say, worlds ω_3 and ω_4 than I am in worlds ω_1 and ω_2, my expected inaccuracy for Rita's distribution falls near the higher end of the values in the fourth column of the table.

In general, if an agent employs the scoring rule I to measure inaccuracy, the agent's credence distribution is cr, and the finite set of worlds under consideration is $\{\omega_1, \omega_2, \dots, \omega_n\}$, the agent's expected inaccuracy for any distribution cr' is:

$$\mathrm{EI_{cr}}(\mathrm{cr}') = \mathrm{I}(\mathrm{cr}', \omega_1) \cdot \mathrm{cr}(\omega_1) + \mathrm{I}(\mathrm{cr}', \omega_2) \cdot \mathrm{cr}(\omega_2) + \dots + \mathrm{I}(\mathrm{cr}', \omega_n) \cdot \mathrm{cr}(\omega_n) \tag{10.6}$$

This equation generalizes the expected inaccuracy calculation of Equation (10.5) above. The notation $\mathrm{EI_{cr}}(\mathrm{cr}')$ indicates that we are calculating the expected inaccuracy of credence distribution cr', as judged *from the point of view* of credence distribution cr.[19]

Equation (10.6) allows me to calculate my expected inaccuracy for any credence distribution, probabilistic or otherwise. If I wanted, I could even calculate my expected inaccuracy for my *own* credence distribution. That is, I could calculate $\mathrm{EI_{cr}}(\mathrm{cr})$. But this calculation is fraught. When I calculate my expected inaccuracy for my own current credences and compare it to the inaccuracy I expect for someone else's credences, I might find that I expect that other distribution to be more accurate than my own. We will say that distribution cr' **defeats cr in expectation** if

$$\mathrm{EI}_{\mathrm{cr}}(\mathrm{cr}') < \mathrm{EI}_{\mathrm{cr}}(\mathrm{cr}) \tag{10.7}$$

Your credence distribution defeats mine in expectation when, from the point of view of my own credence distribution, I expect yours to be less inaccurate than mine.

Being defeated in accuracy expectation is not quite as bad as being accuracy-dominated. Being defeated in expectation is kind of like having a twin sister who takes all the same classes as you but has a better GPA. (Being accuracy-dominated is like your twin's getting a better grade than you *in every single class*.) Still, being defeated in expectation is a rational flaw. Joyce writes:

> If, relative to a person's own credences, some alternative system of beliefs has a lower expected epistemic [inaccuracy], then, by her own estimation, that system is preferable from the epistemic perspective. This puts her in an untenable doxastic situation. She has a *prima facie* epistemic reason, grounded in her beliefs, to think that she should not be relying on those very beliefs. This is a probabilistic version of Moore's paradox. Just as a rational person cannot fully believe "*X* but I don't believe *X*," so a person cannot rationally hold a set of credences that require her to estimate that some other set has higher epistemic utility. [Such a] person is... in this pathological position: her beliefs undermine themselves. (2009, p. 277)

The idea that rational agents avoid being defeated in expectation is related to our earlier weather-forecaster discussion of stability and credence elicitation. Lewis (1971) calls a distribution that assigns itself the highest expected accuracy **immodest**. ("When asked which method has the best estimated accuracy, the immodest method answers: 'I have'.") He then relates immodesty to an agent's epistemic goals:

> If you wish to maximize accuracy in choosing a [credence-assignment] method, and you have knowingly given your trust to any but an immodest method, how can you justify staying with the method you have chosen? If you really trust your method, and you really want to maximize accuracy, you should take your method's advice and maximize accuracy by switching to some other method that your original method recommends. If that method is also not immodest, and you trust it, and you still want to maximize accuracy, you should switch again; and so on, unless you happen to hit upon an immodest method. Immodesty is a condition of adequacy because it is a necessary condition for stable trust. (1971, p. 62)

These arguments from Joyce and Lewis support the following principle:

Admissibles Not Defeated: If an agent's credence distribution is rationally permissible, and she measures inaccuracy with an acceptable scoring rule, then there will not exist any distribution that she expects to be more accurate than her own.

Admissibles Not Defeated says that under an acceptable scoring rule, no credence distribution that is rationally permissible will take itself to be defeated in expectation by another distribution.[20]

Admissibles Not Defeated relates two elements: a credence distribution and a scoring rule. If we've already settled on an acceptable scoring rule, we can use Admissibles Not Defeated to test the rational permissibility of a credence distribution. But we can also argue in the other direction: If we know a particular credence distribution is rational, we can use Admissibles Not Defeated to argue that particular scoring rules are not acceptable.

For example, suppose I'm certain a fair die has just been rolled, but I know nothing about the outcome. I entertain six propositions, one for each possible outcome of the roll, and let's imagine that I assign each of those propositions a credence of 1/6. That is, my credence distribution cr assigns $\mathrm{cr}(1) = \mathrm{cr}(2) = \mathrm{cr}(3) = \mathrm{cr}(4) = \mathrm{cr}(5) = \mathrm{cr}(6) = 1/6$. This seems at least a rationally *permissible* distribution in my situation.

But now suppose that, in addition to having this perfectly permissible credence distribution, I also use the absolute-value scoring rule to assess accuracy. I entertain six possible worlds—call them ω_1 through ω_6, with the subscripts indicating how the roll comes out in a given world. In world ω_1, the roll comes out one, so $\mathrm{tv}_{\omega_1}(1) = 1$ while the tv_{ω_1}-value of each of the other outcomes is zero. Thus we have

$$\mathrm{I_{ABS}}(\mathrm{cr}, \omega_1) = |1 - 1/6| + 5 \cdot |0 - 1/6| = 10/6 = 5/3 \qquad (10.8)$$

A bit of reflection will show that $\mathrm{I_{ABS}}(\mathrm{cr}, \omega_2)$ through $\mathrm{I_{ABS}}(\mathrm{cr}, \omega_6)$ also equal 5/3, for similar reasons. Recalling that I assign credence 1/6 to each of the six possible worlds, my expected inaccuracy for my own credence distribution is

$$\mathrm{EI_{cr}}(\mathrm{cr}) = 6 \cdot (5/3 \cdot 1/6) = 5/3 \qquad (10.9)$$

Next I consider my crazy friend Ned, who has the same evidence as me but assigns credence 0 to each of the six roll-outcome propositions. That is, Ned's distribution cr_N assigns $cr_N(1) = cr_N(2) = cr_N(3) = cr_N(4) = cr_N(5) = cr_N(6) = 0$. How inaccurate do I expect Ned to be? Again, in ω_1, $tv_{\omega_1}(1) = 1$ while the tv_{ω_1}-value of each other outcome is 0. So

$$I_{ABS}(cr_N, \omega_1) = |1 - 0| + 5 \cdot |0 - 0| = 1 \tag{10.10}$$

Similar calculations show that, as measured by the absolute-value score, in each possible world Ned's distribution will have an inaccuracy of 1. When I calculate my expected inaccuracy for Ned, I get

$$EI_{cr}(cr_N) = 6 \cdot (1 \cdot 1/6) = 1 \tag{10.11}$$

And now we run into a problem: 1 is less than 5/3. If I calculate inaccuracy using the absolute-value rule, I will expect Ned's distribution to be less inaccurate than my own; my credence distribution is defeated in expectation by Ned's. Yet Ned's distribution isn't *better* than mine in any epistemic sense—in fact, the Principal Principle would say that my distribution is rationally required while his is rationally forbidden! Something has gone wrong, and it isn't the credences I assigned. Instead, it's the scoring rule I used to compare my credences with Ned's.

We can use this example to construct an argument against the absolute-value score as an acceptable scoring rule. In the example, my credence distribution is rationally permissible. According to Admissibles Not Defeated, a rationally permissible distribution cannot be defeated in expectation on any acceptable scoring rule. On the absolute-value rule, my credence distribution *is* defeated in expectation (by Ned's). So the absolute-value scoring rule is not an acceptable inaccuracy measure. (This argument is similar to an argument we made against calibration as an accuracy measure, on the grounds that calibration made perfectly rational forecaster credences look unstable and therefore irrational.)

The Ned example cannot be used to make a similar argument against the Brier score. Exercise 10.4 shows that if I had used the Brier score, I would have expected my own credence distribution to be more accurate than Ned's. In fact, the Brier score is an example of a proper scoring rule:

Proper Scoring Rule: A scoring rule is proper just in case any agent with a probabilistic credence distribution who uses that rule takes herself to defeat in expectation every other distribution over the same set of propositions.

The absolute-value scoring rule is not proper. The Brier score is: a probabilistic agent who uses the Brier score will always expect herself to do *better* with respect to accuracy than any other distribution she considers.[21] The Brier score is not the only scoring rule with this feature. For the sake of illustration, here's another proper scoring rule:[22]

$$\begin{aligned} \mathrm{I}_{\mathrm{LOG}}(\mathrm{cr}, \omega) = {} & [-\log(1 - |\mathrm{tv}_\omega(X_1) - \mathrm{cr}(X_1)|)] + \ldots \\ & + [-\log(1 - |\mathrm{tv}_\omega(X_n) - \mathrm{cr}(X_n)|)] \end{aligned} \tag{10.12}$$

Historically, the Brier score has been favored over the absolute-value score for inaccuracy measurement because Brier is a proper scoring rule.[23] Of course, propriety gives us no means of choosing between the Brier score and other proper scores such as the logarithmic rule of Equation (10.12). But it turns out we don't need to. Predd et al. (2009) showed that a Gradational Accuracy Theorem can be proven for *any* separable, proper scoring rule (not just the Brier score). So, for instance, on the logarithmic scoring rule any non-probabilistic credence distribution will be accuracy-dominated by some probabilistic distribution over the same propositions, but no probabilistic distribution will be dominated. The same is not true for the absolute-value score. In fact, if you look back to the Crazy Ned example, you'll find that Crazy Ned's non-probabilistic distribution accuracy-dominates my probabilistic distribution cr. In each of the six possible worlds, $\mathrm{I}_{\mathrm{ABS}}(\mathrm{cr}_N, \omega) = 1$ while $\mathrm{I}_{\mathrm{ABS}}(\mathrm{cr}, \omega) = 5/3$. On an improper scoring rule, a non-probabilistic distribution may accuracy-dominate a probabilistic one.

10.3.3 Are improper rules unacceptable?

We now have a clear argumentative path to probabilism. Suppose we establish that all acceptable scoring rules are proper. Then, regardless of any further distinctions we might make among the proper rules, it will turn out that all nonprobabilistic credence distributions can be accuracy-dominated, while no probabilistic distributions can be. Given Admissibles Not Dominated, we will be able to establish that credence distributions violating the probability axioms are irrational. So can we use propriety as an adequacy criterion for scoring rules?

A proper scoring rule is one on which probabilistic distributions always expect themselves to be more accurate than the alternatives. But why focus on what *probabilistic* distributions expect? Inaccuracy measurement has many applications, and in many of those applications (including one we'll see in Section 10.5) it is already assumed that probabilistic credence distributions are rational. In such situations we want an accuracy measure that interacts well with probabilistic distributions, so proper scoring rules are a natural fit, and it's traditional to apply the Brier score because of its propriety. But when an inaccuracy measure is used to *argue* for probabilism—as in the Gradational Accuracy Argument—it seems question-begging to privilege probabilistic distributions in selecting that scoring rule. For instance, our Crazy Ned argument against the absolute-value score *started* by assuming that my probabilistic distribution assigning credence 1/6 to each possible die-roll outcome was rationally permissible. We then criticized the absolute-value score on the grounds that it made that distribution look unstable and therefore irrational. Yet this criticism looks circular in the course of a debate about the rational status of credences satisfying the probability axioms.

In his (2009), Joyce moved from his old reasons for favoring the Brier score to a new approach that explicitly begins with the rational permissibility of probabilistic distributions. While I won't go into the specifics of that argument here, it takes as a premise that given any numerical distribution satisfying the probability axioms, there exists some situation in which it would be rationally permissible for an agent to assign those values as her credences. Admittedly, this premise—that probabilistic credences are rationally *permitted*—is weaker than the ultimate conclusion of Joyce's accuracy-dominance argument—that probabilistic credences are rationally *required*. Still, without any independent support for the premise, it feels like we're assuming the rationality of probabilistic credences in order to prove the rationality of probabilistic credences. It sounds like Linearity In, Linearity Out to me.[24]

Joyce does try to provide independent support for his premise. He argues that for any probabilistic distribution, we could imagine a situation in which an agent is rationally certain that those values reflect the objective chances of the propositions in question. By the Principal Principle, the agent would then be rationally required to assign the relevant values as her credences.

Yet recall our characters Mr. Prob, Mr. Bold, and Mr. Weak. Mr. Prob satisfies the probability axioms, while Mr. Bold violates Finite Additivity by having his credence in each proposition be the square-root of Mr. Prob's credence in that proposition. Mr. Bold happily assigns a higher credence to every uncertain proposition than Mr. Prob does. In arguing for probabilism, we look to establish that Mr. Bold's (and Mr. Weak's) credences are rationally forbidden. If we could establish that rational credences must match the numerical values of known frequencies or objective chances, then in many situations

Mr. Bold's distribution could be ruled out immediately, because frequencies and chances must each be additive.[25] But part of Mr. Bold's boldness is that even when he and Mr. Prob are both certain that a particular proposition has a particular nonextreme chance, he's willing to assign that proposition a higher credence than its chance value. Mr. Bold is willing to be more confident of a given experimental outcome than its numerical chance!

What if, when confronted with a fair die roll like the one in the Crazy Ned example, Mr. Bold maintains that it is rationally *impermissible* to assign a credence of 1/6 to each outcome? It's not that Mr. Bold disagrees with us about what the chances are; it's that he disagrees with us about whether rationally permissible credences equal the chances.[26] Faced with this position, our argument against the absolute-value score could not get off the ground, and would not favor the Brier score over absolute-value in constructing a Gradational Accuracy Argument. Similarly, Joyce's argument for his premise would go nowhere, because Mr. Bold clearly rejects the Principal Principle.[27] While we might intuitively feel like Mr. Bold's position is crazy, the accuracy-based *arguments* against it are question-begging.

10.4 Do we really need Finite Additivity?

Let's step back and take a broader view of the arguments discussed so far in this chapter. Some authors don't think accuracy considerations are central to assessing doxastic attitudes for rationality. But among those who embrace an accuracy-based approach, a few principles are uncontroversial. Everyone accepts Admissibles Not Dominated, and most authors seem okay with Admissibles Not Defeated. Everyone thinks accuracy measures should be truth-directed, and most are on board with separability. Controversy arises when we try to put more substantive constraints on the set of acceptable scoring rules. In order to run a gradational accuracy argument for probabilism, we need to narrow the acceptable scoring rules to the set of proper scores (or one of the other restricted sets Joyce considers in his 1998 and 1999). But arguments for such a restricted set often turn out to be question-begging.

What if we didn't try to narrow the set so far—what if we worked only with constraints on scoring rules that are entirely uncontroversial? In Exercise 10.3, you'll show that as long as one's scoring rule is truth-directed, Admissibles Not Dominated endorses Normality and Non-Negativity as rational constraints on credence. As usual, Finite Additivity is the most difficult Kolmogorov axiom to establish. But an excellent (1982) paper by Dennis Lindley shows how close we can get to full probabilism without strong constraints on our scoring rules.

Lindley assumes Admissibles Not Dominated, then lays down some very minimal constraints on acceptable scoring rules. I won't work through the details, but besides separability and Truth-Directedness he assumes (for instance) that an acceptable scoring rule must be smooth—your score doesn't suddenly jump when you slightly increase or decrease your credence in a proposition. Lindley shows that these thin constraints on scoring rules suffice to narrow down the class of rationally permissible credence distributions, and narrow it down more than just Normality and Non-Negativity would. In fact, every rationally permissible credence distribution is either probabilistic (it satisfies all three Kolmogorov axioms) or can be altered by a simple transformation into a probabilistic distribution. The permissible credence distributions stand to the probabilistic ones in something like the relation Mr. Bold and Mr. Weak stand to Mr. Prob. Mr. Prob satisfies Finite Additivity; Mr. Bold and Mr. Weak don't; but their credences can be converted into Mr. Prob's by a simple mathematical operation (squaring for Mr. Bold; square-rooting for Mr. Weak).[28]

How should we interpret this result? One reading would be that, as long as we rely exclusively on accuracy arguments, we will have to grant that Mr. Bold and Mr. Weak, despite not satisfying Finite Additivity, have doxastic attitudes just as rational as Mr. Prob's. Rational requirements on credences are stronger than just Normality and Non-Negativity—Lindley's result isn't anything goes, and he describes some distributions that satisfy those two constraints but are *not* transformable in the relevant manner into probabilities. But the requirements we get from Lindley are not as strong as Finite Additivity, and do not rule out Mr. Weak or Mr. Bold. So perhaps those characters are perfectly rational, and Finite Additivity is not a rational requirement.

A second reading, however, would be that Mr. Prob, Mr. Bold, and Mr. Weak aren't really distinguishable characters; they don't actually have differing doxastic attitudes in any significant sense. We have stipulated that upon hearing a fair coin was flipped, Mr. Prob assigns credence 1/2 to heads, while Mr. Bold assigns $1/\sqrt{2} \approx .707$. But do they have importantly different outlooks on the world? Their distributions are ordinally equivalent—Mr. Prob is more confident of X than Y just in case Mr. Bold is as well. And both of them satisfy certain structural constraints, such as Normality, Non-Negativity, and our credal Entailment rule. In real life these characters would think and act in many of the same ways; a functionalist might argue that their doxastic attitudes are identical.

Perhaps Mr. Bold stands to Mr. Prob in much the same relation that Fahrenheit measurements of temperature stand to Celsius. Representing temperatures as numbers requires us to introduce a measurement regime, which necessitates some arbitrary choice: what numbers should we assign to water's freezing point, to its boiling point, etc.? Whichever choices we make, the same underlying kinetic phenomena are portrayed—they may contain different *numbers*, but measurements of 0°C and 32°F describe the same state of the world. Perhaps, instead of there being two different characters Mr. Prob and Mr. Bold, the numerical credence distributions we've associated with these characters are just two different representations of the same underlying attitudes, utilizing two different measurement regimes. The relation between measurement schemes wouldn't be quite as straightforward as that between Celsius and Fahrenheit; something more than an affine transformation is involved. But from a mathematical point of view, getting from Mr. Prob's distribution to Mr. Bold's is a simple affair. Lindley identifies a whole family of numerical distributions that are simply transformable into probabilities, and Mr. Bold's is among them.

On this reading, Lindley's result establishes Finite Additivity as a rational requirement in the only way that could possibly matter. A rational agent's credences may be depicted by any one of a number of interrelated numerical distributions, depending on the measurement conventions of the person doing the depicting. To say that Finite Additivity is rationally required is to say that *at least one* of these distributions satisfies it; it's to say that a rational agent's attitudes are *representable* as additive, even if non-additive representations are available as well. Lindley shows that, given minimal conditions on an acceptable accuracy score, every admissible credence distribution either satisfies Finite Additivity or can be transformed into a distribution that does. And there's nothing more substantive than this to the claim that rationality requires Finite Additivity.[29]

To argue against this reading, one would have to argue that there can be significant, cognitive differences between an individual with Mr. Prob's credences and one with Mr. Bold's. If that were the case, then the difference between particular probabilistic and non-probabilistic distributions would not come down to just a choice among measurement schemes. We would be able to find individuals in the world who revealed through their thought, talk, or action that they were like Mr. Bold but not Mr. Prob, and we could have a meaningful conversation about whether their violation of Finite Additivity revealed them to be irrational.

In Chapter 1, I motivated the move from comparative to quantitative confidence models by noting that agents with ordinally equivalent opinions may nevertheless disagree on the relative sizes of confidence *gaps*. Given a tautology, a contradiction, and the proposition that a fair coin came up heads, Mr. Prob and Mr. Bold will rank these three propositions in the same order with respect to confidence. But if we asked an agent like Mr. Prob, he might say that he is more confident in heads than in the contradiction *by the same amount* that he is more confident in the tautology than in heads. A Bold-type wouldn't say that. (Mr. Bold has a larger gap between heads and the contradiction than he has between heads and the tautology.) These sorts of conversations do happen in the real world; perhaps they establish the doxastic differences we seek. Yet I worry about basing our case on agents' self-reports of their psychological states, which are notoriously unreliable. And I worry especially about relying upon conversations we've observed in a culture like ours, which teaches people to measure confidence on something like a linear percentage scale from a very young age.

In Part III of this book, I suggested we assess Bayesian epistemology by considering its applications; I focused especially on applications to confirmation and decision theory. Differences in confidence gaps between ordinally equivalent credence distributions may be highly significant when it comes to decision theory. If I am offered a gamble that yields a small profit on P but a major loss on $\sim P$, my decision will depend not only on whether I find P more likely than $\sim P$, but also on *how much more likely* I find it. So practical rationality may make confidence gap sizes observable in behavior.

Yet we saw in Chapter 8 that the differences between Mr. Prob's and Mr. Bold's credence distributions can be practically neutralized if those agents apply different valuation functions. If Mr. Prob combines his credences and utilities to generate preferences by maximizing expected value, and Mr. Bold combines his credences and identical utilities to generate preferences using a different function, Mr. Prob and Mr. Bold will wind up with the same preferences among acts. In that case, the numerical differences—including confidence-gap differences—between Mr. Prob's and Mr. Bold's credences will make no difference to how they behave. Moreover, Mr. Prob and Mr. Bold will both satisfy the preference axioms that make decision theory's account of practical rationality appealing. So decision theory seems perfectly compatible with reading Mr. Prob and Mr. Bold as just different representations of the same acting individual.

The situation seems to me much more open-ended when it comes to confirmation theory. As with decision theory, confirmation results depend not just on confidence rankings but also on quantitative relations among numerical credence values. In Section 6.4.2 we investigated credence distributions relative to which observing a black raven more strongly confirms the hypothesis that all ravens are black than does observing a non-black, non-raven. The Bayesian solution to the Ravens Paradox presented there describes two conditions on such distributions (Equations (6.10) and (6.11)). The second of those conditions is about the sizes of gaps—it asks whether learning a particular hypothesis would change how much more confident you were in one proposition than another. Despite their ordinal agreements, characters like Mr. Prob and Mr. Bold have different ratios between their credences in particular propositions. So Equation (6.11) might be satisfied by one of them but not by the other. This means that if Mr. Prob and Mr. Bold apply traditional Bayesian confirmation measures, they may disagree on whether the ravens hypothesis is more strongly confirmed by a black raven or by a red herring, which seems like a genuine difference in attitudes.[30] Confirmation is one of many non-decision-theoretic applications of Bayesian epistemology (coherence of a belief set, measuring information content, etc.) where it seems like confidence-gap sizes might make a real difference.

Perhaps in each of those applications we could play a trick similar to the one we used in decision theory. In decision theory we compensated for Mr. Bold's non-additive credence distribution by having him use a non-standard valuation function; the combination yielded act preferences identical to Mr. Prob's. What happens if Mr. Bold also uses a non-traditional confirmation measure? Perhaps there's an odd-looking confirmation measure Mr. Bold could apply which, despite Mr. Bold's credence differences with Mr. Prob, would leave the two agents with identical judgments about confirmational matters.[31] It's unclear, though, how such a non-traditional measure would stand up to the arguments, intuitive considerations, and adequacy conditions that have been deployed in the debate over confirmation measures. I know of no literature on this subject.

Where does that leave Finite Additivity as a rational constraint? As it stands, I think that applications of Bayesianism to theoretical rationality (how we infer, how we reason, how we determine what supports what) have a better chance of drawing real contrasts between Mr. Prob and Mr. Bold than practical applications do. It's also worth noting that Chapter 6's confirmation-theoretic results rely heavily on credence distributions' actually satisfying Finite Additivity. So it may turn out that an appealing account of agents' theoretical judgments will assess those judgments as rational only if the agent's attitudes are genuinely probabilistic. But that is pure speculation on my part.

10.5 An accuracy argument for Conditionalization

Up to this point we've considered accuracy-based arguments for only *synchronic* Bayesian norms. We've found that establishing probabilism on non-circular grounds is somewhat difficult. But if you've already accepted probabilism, a remarkable accuracy-based argument for updating by Conditionalization becomes available. The relevant result was proven by Hilary Greaves and David Wallace (2006).[32] We begin by restricting our attention to proper scoring rules. Doing so is non-circular in this context, because we imagine that we've already accepted probabilism as rationally required. This allows us to appeal to the fact that proper scores are credence-eliciting for probabilistic credences as a reason to prefer them.

Greaves and Wallace think of Conditionalization as a *plan* one could adopt for how to change one's credences in response to one's future evidence. Imagine we have an agent at time t_i with probabilistic credence distribution cr_i, who is certain she will gain some evidence before t_j. Imagine also that there's a finite partition of propositions $\{E_1, E_2, \ldots, E_n\}$ in $\mathcal{L}$ such that the agent is certain the evidence gained will be a member of that partition. The agent can then form a plan for how she intends to update—she says to herself, "If I get evidence E_1, I'll update my credences to such-and-such"; "If I get evidence E_2, I'll update my credences to so-and-so"; etc. In other words, an updating plan is a function from members of the evidence partition to cr_j distributions she would assign were she to receive that evidence. Conditionalization is the plan that sets $\text{cr}_j(\cdot) = \text{cr}_i(\cdot \mid E_m)$ in response to learning E_m between t_i and t_j.

Next, Greaves and Wallace show how, given a particular updating plan, the agent can calculate from her point of view at t_i an expectation for how inaccurate that plan will be.[33] Roughly, the idea is to figure out what credence distribution the plan would generate in each possible world, measure how inaccurate that distribution would be in that world, multiply by the agent's t_i confidence in that possible world, then sum the results. More precisely, the expectation calculation proceeds in six steps:

1. Pick a possible world ω to which the agent assigns nonzero credence at t_i.
2. Figure out which member of the partition $\{E_1, E_2, \ldots, E_n\}$ the agent will receive as evidence between t_i and t_j if ω turns out to be the actual world. (Because possible worlds are maximally specified, there will always be a unique answer to this question.) We'll call that piece of evidence E.
3. Take the updating plan being evaluated and figure out what credence distribution it recommends to the agent if she receives evidence E

between t_i and t_j. This is the credence distribution the agent will assign at t_j if ω is the actual world and she follows the plan in question. We'll call that distribution cr_j.

4. Whichever scoring rule we've chosen (among the proper scoring rules), use it to determine the inaccuracy of cr_j if ω is the actual world. (In other words, calculate $\mathrm{I}(\mathrm{cr}_j, \omega)$.)
5. Multiply that inaccuracy value by the agent's t_i credence that ω is the actual world. (In other words, calculate $\mathrm{I}(\mathrm{cr}_j, \omega) \cdot \mathrm{cr}_i(\omega)$.)
6. Repeat this process for each world to which the agent assigns positive credence at t_i, then sum the results.

This calculation has the t_i agent evaluate an updating plan by determining what cr_j distribution that plan would recommend in each possible world. She assesses the recommended distibution's accuracy in that world, weighting the result by her confidence that the world in question will obtain. Repeating this process for each possible world and summing the results, she develops an overall expectation of how accurate her t_j credences will be if she implements the plan.

Greaves and Wallace go on to prove the following theorem:

Accuracy Updating Theorem: For any proper scoring rule, probabilistic distribution cr_i, and evidential partition in $\mathcal{L}$, a t_i agent who calculates expected inaccuracies as described above will find Conditionalization more accurate than any updating plan that diverges from it.

The Accuracy Updating Theorem demonstrates that from her vantage point at t_i, an agent with probabilistic credences and a proper scoring rule will expect to be most accurate at t_j if she updates by Conditionalization. Given a principle something like Admissibles Not Defeated for updating plans, we can use this result to argue that no updating plan deviating from Conditionalization is rationally acceptable.

Does this argument show that the agent is rationally required to update by Conditionalization between t_i and t_j? If she's interested in minimizing expected inaccuracy, then at t_i she should certainly *plan* to update by conditionalizing—of all the updating plans available to the agent at t_i, she expects Conditionalization to be most accurate. Yet being required to make a plan is different from being required to implement it. Even if the agent remembers at t_j what she planned at t_i, why should the t_j agent do what her

t_i self thought best? Among other things, the t_j agent has more evidence than her t_i self did.

This is the same problem we identified in Chapter 9 for diachronic Dutch Strategy arguments. The Accuracy Updating Theorem establishes a *synchronic* point about which policy a t_i agent concerned with accuracy will hope her t_j self applies.[34] But absent a substantive premise that agents are rationally required later on to honor their earlier plans, we cannot move from this *synchronic* point to a genuinely *diachronic* norm like Conditionalization.

10.6 Exercises

Problem 10.1. On each of ten consecutive mornings, a weather forecaster reports her credence that it will rain that day. Below is a record of the credences she reported and whether it rained that day:

Day	1	2	3	4	5	6	7	8	9	10
cr(rain)	1/2	1/4	1/3	1/3	1/2	1/4	1/3	1	1/2	1/4
Rain?	Y	N	N	N	Y	Y	N	Y	N	N

Unfortunately, the forecaster's reports turned out not to be perfectly calibrated over this ten-day span. But now imagine she is given the opportunity to go back and change two of the credences she reported over those ten days.[35] What two changes should she make so that her reports over the span become perfectly calibrated? (Assume that changing her credence report does not change whether it rains on a given day.)

Problem 10.2. Throughout this problem, assume the Brier score is used to measure inaccuracy.

(a) Suppose we have an agent who assigns credences to two propositions, X and Y, and those credences are between 0 and 1 (inclusive). Draw a box diagram (like those in Figures 10.2, 10.3, and 10.4) illustrating the possible distributions she might assign over these two propositions. Then shade in the parts of the box in which $\mathrm{cr}(X) \geq \mathrm{cr}(Y)$.

(b) Now suppose that $Y \vDash X$. Use your diagram from part (a) to show that if an agent's credence distribution violates the Entailment rule by assigning $\mathrm{cr}(Y) > \mathrm{cr}(X)$, there will exist a distribution distinct from hers that is more accurate than hers in every logically possible world. (Hint: When $Y \vDash X$, only three of the four corners of your box represent logically possible worlds.)

(c) In Exercise 9.2 we encountered Roxanne, who assigns the following credences (among others) at a given time:

$$\mathrm{cr}(A \,\&\, B) = 0.5 \qquad\qquad \mathrm{cr}(A) = 0.1$$

Construct an alternate credence distribution over these two propositions that is more accurate than Roxanne's in every logically possible world. (Hint: Let $A \,\&\, B$ play the role of proposition Y, and A play the role of X.) To demonstrate that you've succeeded, calculate Roxanne's inaccuracy and the alternate distribution's inaccuracy in each of the three available possible worlds.

Problem 10.3. Assuming only that our inaccuracy scoring rule is truth-directed, argue for each of the following from Admissibles Not Dominated:

(a) Non-Negativity
(b) Normality

Problem 10.4. Return to the Crazy Ned example of Section 10.3.2, in which you assign 1/6 credence to each of the six possible die roll outcomes while Ned assigns each a credence of 0. This time we'll use the Brier score (rather than the absolute-value score) to measure inaccuracy in this example.

(a) Calculate the inaccuracy of your credence distribution in a world in which the die comes up one. Then calculate Ned's inaccuracy in that world.
(b) Calculate your expected inaccuracy for your own distribution, then calculate *your* expected inaccuracy for *Ned's* distribution.
(c) How do your results illustrate the fact that the Brier score is a proper scoring rule?

Problem 10.5. Suppose that at t_i an agent assigns credences to exactly four propositions, as follows:

proposition	cr_i
$P \,\&\, Q$	0.1
$P \,\&\, {\sim}Q$	0.2
${\sim}P \,\&\, Q$	0.3
${\sim}P \,\&\, {\sim}Q$	0.4

The agent is certain that between t_i and t_j, she will learn whether Q is true or false.

(a) Imagine the agent has a very bizarre updating plan: No matter what she learns between t_i and t_j, she will assign the exact same credences to the four propositions at t_j that she did at t_i. Using the six-step process described in Section 10.5, and the Brier score to measure inaccuracy, calculate the agent's expected inaccuracy for this updating plan from her point of view at t_i. (Hint: You only need to consider four possible worlds, one for each of the four possible truth-value assignments to the propositions P and Q.)

(b) Now imagine instead that the agent's updating plan is to generate her t_j credences by conditionalizing her t_i credences on the information she learns between the two times. Calculate the agent's t_i expected inaccuracy for *this* updating plan (using the Brier score to measure inaccuracy once more).

(c) How do your results illustrate Greaves and Wallace's Accuracy Updating Theorem?

Problem 10.6. In this exercise you will prove a limited version of Greaves and Wallace's Accuracy Updating Theorem. Suppose we have an agent who assigns credences to exactly four propositions, as follows:

proposition	cr_i	cr_j
$P \& Q$	s	w
$P \& \sim Q$	t	x
$\sim P \& Q$	u	y
$\sim P \& \sim Q$	v	z

where cr_i is probabilistic and regular. Suppose also that the agent is certain at t_i that between then and t_j she will learn the truth about whether Q obtains. Finally, assume the agent uses the Brier score to measure inaccuracy.

(a) If the agent updates by Conditionalization and learns Q between t_i and t_j, what will be the values of w, x, y, and z (expressed in terms of s, t, u, and v)?

(b) We will now systematically consider updating plans that diverge from Conditionalization, and show that for each such plan, there exists an alternative plan that the agent expects at t_i to have lower inaccuracy.

To begin, suppose that the agent has an updating plan on which she assigns a nonzero value to either x or z in the event that she learns Q is true. Show that if we calculate expected inaccuracy using the six-step process described in Section 10.5, she will expect this plan to have a higher inaccuracy than the plan that assigns the same w through z values in the event that she learns $\sim Q$, the same w and y values if she learns Q, but assigns 0 to both x and z if she learns Q.

(A similar argument can be made to show that the agent should assign $w = y = 0$ if she learns $\sim Q$.)

(c) Your work in part (b) allows us to restrict our attention to updating plans that assign $x = z = 0$ when Q is learned. Use the Gradational Accuracy Theorem to argue that among such plans, for any plan that has the agent assign a non-probabilistic t_j distribution after learning Q, there exists another plan that has her assign a probabilistic distribution at t_j after learning Q and that she expects to have a lower inaccuracy from her point of view at t_i.

(A similar argument can be made for the agent's learning $\sim Q$.)

(d) Given your results in parts (b) and (c), we may now confine our attention to updating plans that respond to learning Q by assigning a probabilistic t_j distribution with $x = z = 0$. Suppose we hold fixed what such a plan assigns when the agent learns $\sim Q$, and test different possible assignments to w and y when the agent learns Q. Find the values of w and y that minimize the agent's t_i expected inaccuracy for her updating plan.

(Useful algebra fact: A quadratic equation of the form $f(k) = ak^2 + bk + c$ with positive a attains its minimum when $k = \frac{-b}{2a}$.)

(e) How do the results of parts (a) and (d) confirm Greaves and Wallace's point that updating by Conditionalization minimizes expected inaccuracy? (Notice that an argument similar to that of part (d) could be made for a plan that disagrees with Conditionalization on what to do when the agent learns $\sim Q$.)

Problem 10.7. Of the three kinds of arguments for probabilism we've considered in this part of the book—Representation Theorem arguments, Dutch Book arguments, and accuracy-based arguments—do you think any of them succeeds in establishing requirements of rationality? Which type of argument do you find most convincing? Explain your answers.

10.7 Further reading

INTRODUCTIONS AND OVERVIEWS

Richard Pettigrew (2013b). Epistemic Utility and Norms for Credences. *Philosophy Compass* 8, pp. 897–908

Eminently readable introduction to accuracy-based arguments for Bayesian norms and particular arguments for probabilism and Conditionalization.

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

Book-length presentation of the entire accuracy program.

CLASSIC TEXTS

Bas C. van Fraassen (1983). Calibration: A Frequency Justification for Personal Probability. In: *Physics Philosophy and Psychoanalysis*. Ed. by R. Cohen and L. Laudan. Dordrecht: Reidel, pp. 295–319

Abner Shimony (1988). An Adamite Derivation of the Calculus of Probability. In: *Probability and Causality*. Ed. by J.H. Fetzer. Dordrecht: Reidel, pp. 151–61

Classic arguments for probabilism on calibration grounds.

Bruno de Finetti (1974). *Theory of Probability*. Vol. 1. New York: Wiley

Contains de Finetti's proof of the mathematical result underlying Joyce's Gradational Accuracy Theorem.

James M. Joyce (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science* 65, pp. 575–603

Foundational article that first made the accuracy-dominance argument for probabilism.

Hilary Greaves and David Wallace (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind* 115, pp. 607–32

Presents the minimizing-expected-inaccuracy argument for updating by Conditionalization.

EXTENDED DISCUSSION

James M. Joyce (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In: *Degrees of Belief*. Ed. by Franz Huber and Christoph Schmidt-Petri. Vol. 342. Synthese Library. Springer, pp. 263–97

Joyce further discusses the arguments in his earlier accuracy article and various conditions yielding privileged classes of accuracy scores.

Dennis V. Lindley (1982). Scoring Rules and the Inevitability of Probability. *International Statistical Review* 50, pp. 1–26

Paper discussed in Section 10.4 in which Lindley shows that even with very minimal conditions on acceptable accuracy scores, every rationally permissible credence distribution is either probabilistic or can be converted to a probabilistic distribution via a simple transformation.

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

Hannes Leitgeb and Richard Pettigrew (2010b). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science* 77, pp. 236–72

Presents alternative accuracy-based arguments for synchronic and diachronic Bayesian norms.

Kenny Easwaran (2013). Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science* 80, 119–42

Shows how expected inaccuracy minimization can be extended in the infinite case to support such controversial norms as Reflection and Conglomerability.

Hilary Greaves (2013). Epistemic Decision Theory. *Mind* 122, pp. 915–52

Jennifer Carr (2017). Epistemic Utility Theory and the Aim of Belief. *Philosophy and Phenomenological Research* 95, pp. 511–34

Selim Berker (2013). Epistemic Teleology and the Separateness of Propositions. *Philosophical Review* 122, pp. 337–93

These papers criticize the teleological epistemology of accuracy-based arguments for rational constraints.

Notes

1. In Chapter 9 I suggested that rational appraisals concern how things look from the agent's own point of view. (It's important that an agent be able to tell *for herself* that her credences leave her susceptible to a Dutch Book.) An agent is often unable to assess the accuracy of her own beliefs, since she lacks access to the truth-values of the relevant propositions. This makes the a priori aspect of the argument for Belief Consistency crucial—an agent with inconsistent beliefs can see *from her own standpoint* that at least some of those beliefs are false, regardless of what contingent facts she may or may not have at her disposal.
2. Small technical note: In the definition of calibration, we ignore values of x that the distribution doesn't assign to any propositions. Shimin Zhao also pointed out to me that, while we define calibration for any *real* x, an agent who assigns credences over a finite set of propositions can be perfectly calibrated only if all of her credence values are rational numbers!
3. Like so many notions in Bayesian epistemology, the idea of accuracy as calibration was hinted at in Ramsey. In the latter half of his (1931), Ramsey asks what it would be for credences "to be consistent not merely with one another but also with the facts" (p. 93). He later writes, "Granting that [an agent] is going to think always in the same way about all yellow toadstools, we can ask what degree of confidence it would be best for him to have that they are unwholesome. And the answer is that it will in general be best for his degree of belief that a yellow toadstool is unwholesome to be equal to the proportion of yellow toadstools which are in fact unwholesome" (p. 97).
4. There's also been some interesting empirical research on how well-calibrated agents' credences are in the real world. A robust finding is that everyday people tend to be overconfident in their opinions—only, say, 70% of the propositions to which they assign credence 0.9 turn out to be true. (For a literature survey see Lichtenstein, Fischoff, and Phillips 1982.) On the other hand, Murphy and Winkler (1977) found weather forecasters' precipitation predictions to be fairly well calibrated—even before the introduction of computer, satellite, and radar improvements made since the 1970s!
5. This example is taken from Joyce (1998).
6. If you're a Regularity devotee (Section 4.2), you may think the forecaster shouldn't assign absolute certainty to snow—what she sees out the window could be clever Hollywood staging! Setting the forecaster's credence in snow to 1 makes the numbers in this example simpler, but the same point could be made using an example with regular credences.
7. Compare the practice in statistics of treating a proposition as a dichotomous random variable with value 1 if true and 0 if false.
8. Notice that we're keeping the numerical values of the distribution cr constant as we measure inaccuracy relative to different possible worlds. $I_{BR}(\mathrm{cr}, \omega)$ doesn't measure the

inaccuracy in world ω of the credence distribution the agent *would* have in that world. Instead, given a particular credence distribution cr of interest to us, we will use $I_{BR}(cr, \omega)$ to measure how inaccurate *that particular numerical distribution* is relative to each of a number of distinct possible worlds.

9. In this chapter we will apply scoring rules only to credence distributions over finitely many propositions. If you're wondering what happens when infinitely many propositions get involved, see (Kelley ms) for some important results and useful references.
10. Named after George Brier—another meteorologist!—who discussed it in his (1950).
11. Notice also that each X_i contributes equally to the sum $I_{BR}(cr, \omega)$. Thus I_{BR} treats each proposition to which the agent assigns a credence in some sense equally. If you thought it was more important to be accurate about some X_j than others, you might want to insert constants into the sum weighting the $(tv(X_i) - cr(X_i))^2$ terms differently. The main mathematical results of this chapter would go through even with such weightings; this follows from a lemma called "Stability" at Greaves and Wallace (2006, p. 627).
12. The second part of the Gradational Accuracy Theorem stands to the first part much as the Converse Dutch Book Theorem stands to the Dutch Book Theorem (Chapter 9).
13. Strictly speaking there are four world-*types* here, a world being assigned to a type according to the truth-values it gives X and Y. But since all the worlds of a particular type will enter into accuracy calculations in the same way, I will simplify discussion by pretending there is exactly one world in each type.
14. Notice that a similar argument could be made for any cr lying outside the square defined by ω_4, ω_2, ω_3, and ω_1. So this argument also shows how to accuracy-dominate a distribution that violates our Maximum rule.

    One might wonder why we *need* an argument that credence-values below 0 or above 1 are irrational—didn't we stipulate our scale for measuring degrees of belief such that no value could fall outside that range? On some ways of understanding credence, arguments for Non-Negativity are indeed superfluous. But one might define credences purely in terms of their role in generating preferences (as discussed in Chapter 8) or sanctioning bets (as discussed in Chapter 9). In that case, there would be no immediate reason why a credence couldn't take on a negative value.
15. Suppose you assign credences to three propositions X, Y, and Z such that X and Y are mutually exclusive and $Z \dashv\vDash X \vee Y$. We establish X-, Y-, and Z-axes, then notice that only three points in this space represent logically possible worlds: $(0, 0, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$. The distributions in this space satisfying Finite Additivity all lie on the plane passing through those three points. If your credence distribution cr violates Finite Additivity, it will not lie on that plane. We can accuracy-dominate it with distribution cr′ that is the closest point to cr lying on the plane. If you pick any one of the three logically possible worlds (call it ω), it will form a right triangle with cr and cr′, with the segment from cr to ω as the hypotenuse and the segment from cr′ to ω as a leg. That makes cr′ closer than cr to ω.
16. To give the reader a sense of how the second part of the Gradational Accuracy Theorem is proven, I will now argue that no point lying inside the box in Figure 10.4 and on the illustrated diagonal may be accuracy-dominated with respect to worlds ω_2 and ω_3. In other words, I'll show how satisfying Negation wards off accuracy domination (assuming one measures inaccuracy by the Brier score).

Start with distribution cr′ in Figure 10.4, which lies on the diagonal and therefore satisfies Negation. Imagine drawing two circles through cr′, one centered on ω_2 and the other centered on ω_3. To improve upon the accuracy of cr′ in ω_2, one would have to choose a distribution closer to ω_2 than cr′—in other words, a distribution lying inside the circle centered on ω_2. To improve upon the accuracy of cr′ in ω_3, one would have to choose a distribution lying inside the circle centered on ω_3. But since cr′ lies on the line connecting ω_2 and ω_3, those circles are tangent to each other at cr′, so there is no point lying inside *both* circles. Thus no distribution is more accurate than cr′ in both ω_2 and ω_3.

17. The dashed line is like a contour line on a topographical map. There, every point on a given contour line lies at the same altitude. Here, every point on the dashed line has the same level of inaccuracy.
18. Here I'm employing a convention that "cr(ω_1)" is the value cr assigns to the proposition that *X* and *Y* have the truth-values they possess in world ω_1. In other words, cr(ω_1) is the cr-value on the first line of the probability table.
19. Readers familiar with decision theory (perhaps from Chapter 7) may notice that the expected-inaccuracy calculation of Equation (10.6) strongly resembles Savage's formula for calculating expected utilities. Here a "state" is a possible world ω_i that might be actual, an "act" is assigning a particular credence distribution cr′, and an "outcome" is the inaccuracy that results if ω_i is actual and one assigns cr′. Savage's expected utility formula was abandoned by Jeffrey because it yielded implausible results when states and acts were not independent. Might we have a similar concern about Equation (10.6)? What if the act of assigning a particular credence distribution is not independent of the state that a particular one of the ω_i obtains? Should we move to a Jeffrey-style expected inaccuracy calculation, and perhaps from there to some analogue of Causal Decision Theory? As of this writing, this question is only just beginning to be explored in the accuracy literature, in articles such as Greaves (2013) and Konek and Levinstein (2019).
20. Notice that Admissibles Not Defeated entails our earlier principle Admissibles Not Dominated. If distribution cr′ accuracy-dominates distribution cr, it will also have a lower expected inaccuracy than cr from cr's point of view (because it will have a lower inaccuracy in every possible world). So being accuracy-dominated is a particularly bad way of being defeated in expectation. (As in sports, it's bad enough to get defeated, but even worse to get *dominated*.) Admissibles Not Defeated says that permissible credence distributions are never defeated in expectation; this entails that they are also never dominated.
21. On a proper scoring rule, a probabilistic agent will always expect her own accuracy to be better than that of any other distribution. On the absolute-value rule, a probabilistic agent will sometimes expect other distributions to be better than her own. Some scoring rules fall in the middle: on such rules, a probabilistic agent will never expect anyone else to do *better* than herself, but she may find other distributions whose expected accuracy is *tied* with her own. To highlight this case, some authors distinguish "strictly proper" scoring rules from just "proper" ones. On a strictly proper scoring rule a probabilistic agent will never find any other distribution that ties hers for accuracy expectation; a merely proper rule allows such ties. I am using the term "proper" the way these authors use "strictly proper". For an assessment of how the distinction between propriety and

strict propriety interacts with the results of this chapter and with varying notions of accuracy dominance (such as "strong" vs. "weak" accuracy domination), see Schervish, Seidenfeld, and Kadane (2009). For an argument that one's commitments to propriety and strict propriety should stand or fall together, see Campbell-Moore and Levinstein (2021).

22. This rule is intended to be applied only to cr-values between 0 and 1 (inclusive).
23. To better understand the Brier score, we visualize it as the Euclidean distance between two points in space. Strictly speaking, though, Euclidean distance is the square-root of the Brier score. As long as we make only ordinal comparisons (whether one distribution is more accurate than, or just as accurate as, another distribution in a given world), that square-root doesn't matter. So all the arguments in previous sections (including the arguments that non-probabilistic distributions can be dominated) go through either way. But square-roots can make a difference to expectation calculations. It turns out that while the Brier score is a proper scoring rule, its square-root (the Euclidean distance) is not.
24. From a Linearity-In, Linearity-Out point of view, Joyce's (2009) argument does have one advantage over attempts to favor the Brier score using propriety considerations. If you're truly worried about making linearity assumptions in the process of establishing probabilism, you might be concerned that Admissibles Not Defeated centers around linear *expectations* of inaccuracy. Joyce's (2009) argument runs from his premise to probabilism using only Admissibles Not Dominated along the way, without invoking Admissibles Not Defeated at all.
25. See note 4 in Chapter 5.
26. Compare Fine (1973, Sect. IIID).
27. See Hájek (2009a) for a very different kind of objection to Joyce's argument.
28. It's worth comparing Lindley's result to Cox's Theorem (though the latter does not invoke considerations of accuracy). Richard Cox (1946, 1961) laid down a set of minimal constraints on an agent's credence distribution, such as: the agent assigns equal credences to logically equivalent propositions; the agent's credence in $\sim P$ is a function of her credence in P; her credence in $P \& Q$ is a function of her credence in Q and her credence in P given Q; the latter function is twice differentiable; etc. He then showed that any credence distribution satisfying these constraints is isomorphic to a probabilistic distribution. For discussion of the mathematics, and of various philosophical concerns about Cox's conditions, see Paris (1994), Halpern (1999), Van Horn (2003), and Colyvan (2004).
29. I'm inclined to read Lindley's own interpretation of his result along these lines. For one thing, Lindley titles his paper "Scoring Rules and the Inevitability of Probability". For another, after noting on page 8 that Admissibles Not Defeated is a kind of Pareto optimality rule, he writes that an agent who chooses any of the distributions permitted by that rule and a minimally acceptable scoring rule is thereby "effectively introducing probabilities".
30. The same goes for Bayesian results mentioned in Chapter 6, note 39 showing that a red herring cannot confirm the ravens hypothesis to anything more than an exceedingly weak degree. These results depend on particular credal differences and ratios being "minute" in absolute terms, so they might go through for Mr. Prob but not for Mr. Bold (or vice versa).

31. Since Mr. Bold's credences are the square-root of Mr. Prob's, an obvious move would be to take whatever confirmation measure Mr. Prob uses and replace all of its credal expressions with their squares.
32. As we'll see, the Greaves and Wallace result focuses on minimizing *expected* inaccuracy. For Conditionalization arguments based on accuracy-domination, see Briggs and Pettigrew (2020) and Williams (ms). For an alternative expected-accuracy approach to updating, see Leitgeb and Pettigrew (2010a,b).
33. It's important that Greaves and Wallace restrict their attention to what they call "available" updating plans. Available plans guide an agent's credal response to her total evidence (including the evidence she imagines she'll receive); they do not allow an agent to set her credences based on further factors not in evidence. For instance, consider the updating plan according to which an agent magically assigns credence 1 to each proposition just in case it's true and credence 0 just in case it's false—even if her evidence isn't fine-grained enough to indicate the truth-values of all the relevant propositions. This would be an excellent plan in terms of minimizing inaccuracy, but it isn't a feasible updating strategy for an agent going forward. This updating plan does not count as "available" in Greaves and Wallace's sense, and so does not compete with Conditionalization for the most accurate updating plan.
34. Like Reflection, the resulting norm is a *synchronic* requirement on an agent's attitudes toward propositions about *diachronic* events.
35. Perhaps via time-machine?

Bibliography of Volumes 1 & 2

Achinstein, Peter (1963). Variety and Analogy in Confirmation Theory. *Philosophy of Science* 3, pp. 207–21.

Adams, Ernest (1962). On Rational Betting Systems. *Archiv für mathematische Logik und Grundlagenforschung* 6, pp. 7–29.

Adams, Ernest (1965). The Logic of Conditionals. *Inquiry* 8, pp. 166–97.

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson (1985). On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *The Journal of Symbolic Logic* 50, pp. 510–30.

Allais, Maurice (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulates et axiomes de l'ecole Américaine. *Econometrica* 21, pp. 503–46.

Amrhein, Valentin, Sander Greenland, and Blake McShane (2019). Scientists Rise up against Statistical Significance. *Nature* 567, pp. 305–7.

Armendt, Brad (1980). Is There a Dutch Book Argument for Probability Kinematics? *Philosophy of Science* 47, pp. 583–8.

Armendt, Brad (1992). Dutch Strategies for Diachronic Rules: When Believers See the Sure Loss Coming. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1, pp. 217–29.

Arntzenius, Frank (1993). The Common Cause Principle. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2, pp. 227–37.

Arntzenius, Frank (2003). Some Problems for Conditionalization and Reflection. *The Journal of Philosophy* 100, pp. 356–70.

Arrow, Kenneth J. (1951). *Social Choice and Individual Values.* New York: John Wiley and Sons.

Bandyopadhyay, Prasanta S. and Malcolm R. Forster (2011). *Philosophy of Statistics.* Vol. 7. Handbook of the Philosophy of Science. Amsterdam: Elsevier.

Barnett, Jeffrey A. (1981). Computational Methods for a Mathematical Theory of Evidence. In: *Proceedings of the 7th International Joint Conference on AI,* pp. 868–75.

Bartha, Paul and Christopher R. Hitchcock (1999). No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes's Theorem. *Philosophy of Science* 66, S339–53.

Berger, James O. and Donald A. Berry (1988). Statistical Analysis and the Illusion of Objectivity. *American Scientist* 76, pp. 159–65.

Bergmann, Merrie, James Moor, and Jack Nelson (2013). *The Logic Book.* 6th edition. New York: McGraw Hill.

Berker, Selim (2013). Epistemic Teleology and the Separateness of Propositions. *Philosophical Review* 122, pp. 337–93.

Bernoulli, Daniel (1738/1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica* 22, pp. 23–36.

Bernoulli, Jacob (1713). *Ars Conjectandi.* Basiliae.

Bertrand, Joseph (1888/1972). *Calcul des probabilités.* 2nd edition. New York: Chelsea Publishing Company.

Bickel, P.J., E.A. Hammel, and J.W. O'Connell (1975). Sex Bias in Graduate Admissions: Data from Berkeley. *Science* 187, pp. 398–404.

Birnbaum, Allan (1962). On the Foundations of Statistical Inference. *Journal of the American Statistical Association* 57, pp. 269–306.

Bolzano, Bernard (1837/1973). *Wissenschaftslehre*. Translated by Jan Berg under the title *Theory of Science*. Dordrecht: Reidel.

Bovens, Luc and Stephan Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Bradley, Darren (2010). Conditionalization and Belief *De Se*. *Dialectica* 64, pp. 247–50.

Bradley, Darren (2011). Self-location Is No Problem for Conditionalization. *Synthese* 182, pp. 393–411.

Bradley, Darren (2015). *A Criticial Introduction to Formal Epistemology*. London: Bloomsbury.

Bratman, Michael E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Brier, George (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review* 78, pp. 1–3.

Briggs, R.A. (2010). Putting a Value on Beauty. In: *Oxford Studies in Epistemology*. Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 3. Oxford University Press, pp. 3–34.

Briggs, R.A. and Richard Pettigrew (2020). An Accuracy-Dominance Argument for Conditionalization. *Noûs* 54, pp. 162–81.

Broome, John (1999). Normative Requirements. *Ratio* 12, pp. 398–419.

Buchak, Lara (2013). *Risk and Rationality*. Oxford: Oxford University Press.

Campbell-Moore, Catrin and Benjamin A. Levinstein (2021). Strict Propriety Is Weak. *Analysis* 81.1, pp. 8–13.

Capotorti, Andrea and Barbara Vantaggi (2000). Axiomatic Characterization of Partial Ordinal Relations. *International Journal of Approximate Reasoning* 24, pp. 207–19.

Carnap, Rudolf (1945). On Inductive Logic. *Philosophy of Science* 12, pp. 72–97.

Carnap, Rudolf (1947). On the Application of Inductive Logic. *Philosophy and Phenomenological Research* 8, pp. 133–48.

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

Carnap, Rudolf (1955/1989). Statistical and Inductive Probability. In: *Readings in the Philosophy of Science*. Ed. by Baruch A. Brody and Richard E. Grandy. 2nd edition. Hoboken: Prentice-Hall, pp. 279–87.

Carnap, Rudolf (1962a). *Logical Foundations of Probability*. 2nd edition. Chicago: University of Chicago Press.

Carnap, Rudolf (1962b). The Aim of Inductive Logic. In: *Logic, Methodology, and the Philosophy of Science*. Ed. by P. Suppes, E. Nagel, and A. Tarski. Stanford University: Stanford University Press, pp. 303–18.

Carnap, Rudolf (1971). A Basic System of Inductive Logic, Part 1. In: *Studies in Inductive Logic and Probability*. Ed. by Rudolf Carnap and Richard C. Jeffrey. Vol. I. Berkeley: University of California Press, pp. 33–166.

Carnap, Rudolf (1980). A Basic System of Inductive Logic, Part 2. In: *Studies in Inductive Logic and Probability*. Ed. by Richard C. Jeffrey. Vol. II. Berkeley: University of California Press, pp. 7–156.

Carr, Jennifer (2017). Epistemic Utility Theory and the Aim of Belief. *Philosophy and Phenomenological Research* 95, pp. 511–34.

Cartwright, Nancy (1979). Causal Laws and Effective Strategies. *Noûs* 13, pp. 419–37.

Carver, Ronald P. (1978). The Case against Statistical Significance Testing. *Harvard Educational Review* 48, pp. 378–99.

Chandler, Jake (2013). Contrastive Confirmation: Some Competing Accounts. *Synthese* 190, pp. 129–38.

Cherniak, Christopher (1986). *Minimal Rationality.* Cambridge, MA: The MIT Press.

Chihara, C. (1981). Quine and the Confirmational Paradoxes. In: *Midwest Studies in Philosophy 6: Foundations of Analytic Philosophy.* Ed. by P. French, H. Wettstein, and T. Uehling. Minneapolis: University of Minnesota Press, pp. 425–52.

Christensen, David (1991). Clever Bookies and Coherent Beliefs. *The Philosophical Review* 100, pp. 229–47.

Christensen, David (1999). Measuring Confirmation. *The Journal of Philosophy* 96, pp. 437–61.

Christensen, David (2001). Preference-Based Arguments for Probabilism. *Philosophy of Science* 68, pp. 356–76.

Christensen, David (2004). *Putting Logic in its Place.* Oxford: Oxford University Press.

Colyvan, Mark (2004). The Philosophical Significance of Cox's Theorem. *International Journal of Approximate Reasoning* 37, pp. 71–85.

Cox, Richard T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics* 14, pp. 1–13.

Cox, Richard T. (1961). *The Algebra of Probable Inference.* Baltimore, MD: The Johns Hopkins Press.

Crupi, Vincenzo, Branden Fitelson, and Katya Tentori (2008). Probability, Confirmation, and the Conjunction Fallacy. *Thinking & Reasoning* 14, pp. 182–99.

Crupi, Vincenzo, Katya Tentori, and Michel Gonzalez (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science* 74, pp. 229–52.

Davidson, Donald (1984). *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press.

Davidson, Donald, J.C.C. McKinsey, and Patrick Suppes (1955). Outlines of a Formal Theory of Value, I. *Philosophy of Science* 22, pp. 14–60.

de Finetti, Bruno (1931/1989). Probabilism: A Critical Essay on the Theory of Probability and the Value of Science. *Erkenntnis* 31, pp. 169–223. Translation of B. de Finetti, *Probabilismo,* Logos 14, pp. 163–219.

de Finetti, Bruno (1937/1964). Foresight: Its Logical Laws, its Subjective Sources. In: *Studies in Subjective Probability.* Ed. by Henry E. Kyburg Jr and H.E. Smokler. New York: Wiley, pp. 94–158. Originally published as "La prévision; ses lois logiques, ses sources subjectives" in *Annales de l'Institut Henri Poincaré* 7, pp. 1–68.

de Finetti, Bruno (1949/1951). La 'Logica del Plausible' Secondo la Conçezione di Polya. *Atti della XLII Riunione, Societa Italiana per il Progresso delle Scienze.* Presented in 1949, published in 1951, pp. 227–36.

de Finetti, Bruno (1974). *Theory of Probability.* Vol. 1. New York: Wiley.

de Finetti, Bruno (1995). *Filosofia della probabilità.* Ed. by Alberto Mura. Milan: Il Saggiatore.

DeGroot, Morris H. (1970). *Optimal Statistical Decisions.* Hoboken, New Jersey: Wiley.

Dempster, Arthur P. (1966). New Methods for Reasoning towards Posterior Distributions Based on Sample Data. *Annals of Mathematical Statistics* 37, pp. 355–74.

Dickson, Michael and Davis Baird (2011). Significance Testing. In: *Philosophy of Statistics.* Ed. by Prasanta S. Bandyopadhyay and Malcolm R. Forster. Vol. 7. Handbook of the Philosophy of Science. Amsterdam: Elsevier.

Domotor, Zoltan (1969). *Probabilistic Relational Structures and their Applications.* Technical Report 144. Stanford University: Institute for Mathematical Studies in the Social Sciences.

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

Easwaran, Kenny (2013). Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science* 80, pp. 119–42.

Easwaran, Kenny (2014a). Decision Theory without Representation Theorems. *Philosophers' Imprint* 14, pp. 1–30.

Easwaran, Kenny (2014b). Regularity and Hyperreal Credences. *Philosophical Review* 123, pp. 1–41.

Eddington, A. (1939). *The Philosophy of Physical Science.* Cambridge: Cambridge University Press.

Edwards, A.W.F. (1972). *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference.* Cambridge: Cambridge University Press.

Eells, Ellery (1982). *Rational Decision and Causality.* Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.

Eells, Ellery (1985). Problems of Old Evidence. *Pacific Philosophical Quarterly* 66, pp. 283–302.

Eells, Ellery and Branden Fitelson (2000). Measuring Confirmation and Evidence. *Journal of Philosophy* 97, pp. 663–72.

Eells, Ellery and Branden Fitelson (2002). Symmetries and Asymmetries in Evidential Support. *Philosophical Studies* 107, pp. 129–42.

Efron, B. (1986). Why Isn't Everyone a Bayesian? *The American Statistician* 40, pp. 1–5.

Egan, Andy (2007). Some Counterexamples to Causal Decision Theory. *Philosophical Review* 116, pp. 93–114.

Egan, Andy and Michael G. Titelbaum (2022). Self-locating Belief. Forthcoming in *The Stanford Encyclopedia of Philosophy.*

Elga, Adam (2000). Self-locating Belief and the Sleeping Beauty Problem. *Analysis* 60, pp. 143–7.

Elga, Adam (2007). Reflection and Disagreement. *Noûs* 41, pp. 478–502.

Elga, Adam (2010). Subjective Probabilities Should Be Sharp. *Philosophers' Imprint* 10, pp. 1–11.

Ellenberg, Jordan (2014). *How Not to Be Wrong: The Power of Mathematical Thinking.* New York: Penguin Press.

Ellis, Robert Leslie (1849). On the Foundations of the Theory of Probabilities. *Transactions of the Cambridge Philosophical Society* VIII, pp. 1–6.

Ellsberg, Daniel (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics* 75, pp. 643–69.

Feldman, Richard (2007). Reasonable Religious Disagreements. In: *Philosophers without Gods: Meditations on Atheism and the Secular Life.* Ed. by Louise M. Antony. Oxford: Oxford University Press, pp. 194–214.

Feller, William (1968). *An Introduction to Probability Theory and its Applications.* 3rd edition. New York: Wiley.

Fermat, Pierre and Blaise Pascal (1654/1929). Fermat and Pascal on Probability. In: *A Source Book in Mathematics.* Ed. by D. Smith. Translated by Vera Sanford. New York: McGraw-Hill, pp. 546–65.

Fine, Terrence L. (1973). *Theories of Probability: An Examination of Foundations.* New York, London: Academic Press.

Fishburn, Peter C. (1981). Subjective Expected Utility: A Review of Normative Theories. *Theory and Decision* 13, pp. 129–99.

Fishburn, Peter C. (1986). The Axioms of Subjective Probability. *Statistical Science* 1, pp. 335–45.

Fisher, Ronald A. (1956). *Statistical Methods and Scientific Inference.* Edinburgh: Oliver and Boyd.

Fitelson, Branden (2006). Logical Foundations of Evidential Support. *Philosophy of Science* 73, pp. 500–12.

Fitelson, Branden (2007). Likelihoodism, Bayesianism, and Relational Confirmation. *Synthese* 156, pp. 473–89.

Fitelson, Branden (2008). A Decision Procedure for Probability Calculus with Applications. *The Review of Symbolic Logic* 1, pp. 111–25.

Fitelson, Branden (2012). Evidence of Evidence Is Not (Necessarily) Evidence. *Analysis* 72, pp. 85–8.

Fitelson, Branden (2015). The Strongest Possible Lewisian Triviality Result. *Thought* 4, pp. 69–74.

Fitelson, Branden and Alan Hájek (2014). Declarations of Independence. *Synthese* 194, pp. 3979–95.

Fitelson, Branden and James Hawthorne (2010a). How Bayesian Confirmation Theory Handles the Paradox of the Ravens. *The Place of Probability in Science.* Ed. by Ellery Eells and J. Fetzer. *Boston Studies in the Philosophy of Science* 284, pp. 247–75.

Fitelson, Branden and James Hawthorne (2010b). The Wason Task(s) and the Paradox of Confirmation. *Philosophical Perspectives* 24, pp. 207–41.

Fitelson, Branden and David McCarthy (ms). Accuracy and Comparative Likelihood. Unpublished manuscript.

Fitelson, Branden and Daniel Osherson (2015). Remarks on "Random Sequences". *Australasian Journal of Logic* 12, pp. 11–16.

Foley, Richard (1987). *The Theory of Epistemic Rationality.* Cambridge, MA: Harvard University Press.

Foley, Richard (1993). *Working without a Net.* Oxford: Oxford University Press.

Foley, Richard (2009). Beliefs, Degrees of Belief, and the Lockean Thesis. In: *Degrees of Belief.* Ed. by Franz Huber and Christoph Schmidt-Petri. Vol. 342. Synthese Library. Springer, pp. 37–48.

Frigerio, Roberta et al. (2005). Education and Occupations Preceding Parkinson's Disease. *Neurology* 65, pp. 1575–83.

Gaifman, H. and M. Snir (1982). Probabilities over Rich Languages. *Journal of Symbolic Logic* 47, pp. 495–548.

Gaifman, Haim (2004). Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements. *Synthese* 140, pp. 97–119.

Galavotti, Maria Carla (2005). *Philosophical Introduction to Probability.* CSLI Lecture Notes 167. Stanford, CA: CSLI Publications.

Gallo, Valentina et al. (2018). Exploring Causality of the Association between Smoking and Parkinson's Disease. *International Journal of Epidemiology* 48, pp. 912–25.

Gandenberger, Greg (2014). *Titelbaum's Counterexample to the Law of Likelihood.* Blog post available at URL: https://gandenberger.org/2014/05/26/titelbaum-counterexample/.

Gandenberger, Greg (2015). A New Proof of the Likelihood Principle. *British Journal for the Philosophy of Science* 66, pp. 475–503.

Gandenberger, Greg (ms). New Responses to Three Purported Counterexamples to the Likelihood Principle. Unpublished manuscript.

Garber, Daniel (1983). Old Evidence and Logical Omniscience in Bayesian Confirmation Theory. In: *Testing Scientific Theories*. Ed. by John Earman. Vol. 10. Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press, pp. 99–132.

Gibbard, A. and W. Harper (1978/1981). Counterfactuals and Two Kinds of Expected Utility. In: *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Ed. by W. Harper, Robert C. Stalnaker, and G. Pearce. Dordrecht: Reidel, pp. 153–90.

Gillies, Donald (2000). Varieties of Propensity. *British Journal for the Philosophy of Science* 51, pp. 807–35.

Glass, David H. and Mark McCartney (2015). A New Argument for the Likelihood Ratio Measure of Confirmation. *Acta Analytica* 30, pp. 59–65.

Glymour, Clark (1980). *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Good, I.J. (1952). Rational Decisions. *Journal of the Royal Statistical Society, Series B* 14, pp. 107–14.

Good, I.J. (1962). Subjective Probability as the Measure of a Non-measurable Set. In: *Logic, Methodology, and the Philosophy of Science*. Ed. by P. Suppes, E. Nagel, and A. Tarski. Stanford, CA: Stanford University Press, pp. 319–29.

Good, I.J. (1967). The White Shoe Is a Red Herring. *British Journal for the Philosophy of Science* 17, p. 322.

Good, I.J. (1968). The White Shoe qua Herring Is Pink. *British Journal for the Philosophy of Science* 19, pp. 156–7.

Good, I.J. (1971). Letter to the Editor. *The American Statistician* 25, pp. 62–3.

Goodman, Nelson (1946). A Query on Confirmation. *The Journal of Philosophy* 43, pp. 383–5.

Goodman, Nelson (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Greaves, Hilary (2013). Epistemic Decision Theory. *Mind* 122, pp. 915–52.

Greaves, Hilary and David Wallace (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind* 115, pp. 607–32.

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

Hacking, Ian (1967). Slightly More Realistic Personal Probability. *Philosophy of Science* 34, pp. 311–25.

Hacking, Ian (1971). The Leibniz-Carnap Program for Inductive Logic. *The Journal of Philosophy* 68, pp. 597–610.

Hacking, Ian (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Hájek, Alan (1996). 'Mises Redux'—Redux: Fifteen Arguments against Finite Frequentism. *Erkenntnis* 45, pp. 209–27.

Hájek, Alan (2003). What Conditional Probability Could Not Be. *Synthese* 137, pp. 273–323.

Hájek, Alan (2009a). Arguments for—or against—Probabilism? In: *Degrees of Belief*. Ed. by Franz Huber and Christoph Schmidt-Petri. Vol. 342. Synthese Library. Springer, pp. 229–51.

Hájek, Alan (2009b). Fifteen Arguments against Hypothetical Frequentism. *Erkenntnis* 70, pp. 211–35.

Hájek, Alan (2011a). Conditional Probability. In: *Philosophy of Statistics*. Ed. by Prasanta S. Bandyopadhyay and Malcolm R. Forster. Vol. 7. Handbook of the Philosophy of Science. Amsterdam: Elsevier, pp. 99–136.

Hájek, Alan (2011b). Triviality Pursuit. *Topoi* 30, pp. 3–15.

Hájek, Alan (2019). Interpretations of Probability. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2019. URL: http://plato.stanford.edu/archives/fall2019/entries/probability-interpret/.

Hájek, Alan and James M. Joyce (2008). Confirmation. In: *The Routledge Companion to Philosophy of Science*. Ed. by Stathis Psillos and Martin Curd. New York: Routledge, pp. 115–28.

Hall, Ned (2004). Two Mistakes about Credence and Chance. *Australasian Journal of Philosophy* 82, pp. 93–111.

Halpern, Joseph Y. (1999). Cox's Theorem Revisited. *Journal of Artificial Intelligence Research* 11, pp. 429–35.

Halpern, Joseph Y. (2003). *Reasoning about Uncertainty*. Cambridge, MA: MIT Press.

Halpern, Joseph Y. (2004). Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. *Proceedings of the Twentieth Conference on Uncertainty in AI*, pp. 226–34.

Halpern, Joseph Y. (2005). Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. In: *Oxford Studies in Epistemology*. Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 1. Oxford: Oxford University Press, pp. 111–42.

Harman, Gilbert (1986). *Change in View*. Boston: The MIT Press.

Harrison-Trainor, Matthew, Wesley H. Holliday, and Thomas F. Icard III (2016). A Note on Cancellation Axioms for Comparative Probability. *Theory and Decision* 80, pp. 159–66.

Hart, Casey and Michael G. Titelbaum (2015). Intuitive Dilation? *Thought* 4, pp. 252–62.

Haverkamp, Nick and Moritz Schulz (2012). A Note on Comparative Probability. *Erkenntnis* 76, pp. 395–402.

Hawthorne, James (2014). Inductive Logic. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2014. URL: http://plato.stanford.edu/archives/win2014/entries/logic-inductive/.

Hawthorne, James (2016). A Logic of Comparative Support: Qualitative Conditional Probability Relations Representable by Popper Functions. In: *The Oxford Handbook of Probability and Philosophy*. Ed. by Alan Hájek and Christopher R. Hitchcock. Oxford: Oxford University Press, pp. 277–95.

Hawthorne, James and Branden Fitelson (2004). Re-solving Irrelevant Conjunction with Probabilistic Independence. *Philosophy of Science* 71, pp. 505–14.

Hedden, Brian (2013). Incoherence without Exploitability. *Noûs* 47, pp. 482–95.

Hempel, Carl G. (1945a). Studies in the Logic of Confirmation (I). *Mind* 54, pp. 1–26.

Hempel, Carl G. (1945b). Studies in the Logic of Confirmation (II). *Mind* 54, pp. 97–121.

Hesse, Mary (1963). *Models and Analogies in Science*. London: Sheed & Ward.

Heukelom, Floris (2015). A History of the Allais Paradox. *The British Journal for the History of Science* 48, pp. 147–69.

Hintikka, Jaakko (1975). Impossible Possible Worlds Vindicated. *Journal of Philosophical Logic* 4, pp. 475–84.

Hitchcock, Christopher R. (2004). Beauty and the Bets. *Synthese* 139, pp. 405–20.

Hitchcock, Christopher R. (2021). Probabilistic Causation. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2021. URL: https://plato.stanford.edu/archives/spr2021/entries/causation-probabilistic/.

Holton, Richard (2014). Intention as a Model for Belief. In: *Rational and Social Agency: The Philosophy of Michael Bratman*. Ed. by Manuel Vargas and Gideon Yaffe. Oxford: Oxford University Press, pp. 12–37.

Hooker, C.A. (1968). Goodman, 'Grue' and Hempel. *Philosophy of Science* 35, pp. 232–47.

Hosiasson-Lindenbaum, Janina (1940). On Confirmation. *Journal of Symbolic Logic* 5, pp. 133–48.

Howson, Colin (1992). Dutch Book Arguments and Consistency. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2, pp. 161–8.

Howson, Colin (2014). Finite Additivity, Another Lottery Paradox and Conditionalisation. *Synthese* 191, pp. 989–1012.

Howson, Colin and Peter Urbach (2006). *Scientific Reasoning: The Bayesian Approach.* 3rd edition. Chicago: Open Court.

Huber, Franz (2016). Formal Representations of Belief. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Spring 2016. URL: https://plato.stanford.edu/archives/spr2016/entries/formal-belief/.

Hume, David (1739–40/1978). *A Treatise of Human Nature.* Ed. by L.A. Selby-Bigge and Peter H. Nidditch. 2nd edition. Oxford: Oxford University Press.

Humphreys, Paul (1985). Why Propensities Cannot Be Probabilities. *Philosophical Review* 94, pp. 557–70.

Icard III, Thomas F. (2016). Pragmatic Considerations on Comparative Probability. *Philosophy of Science* 83, pp. 348–70.

Jackson, Elizabeth G. (2020). The Relationship between Belief and Credence. *Philosophy Compass* 15, pp. 1–13.

Jaynes, E.T. (1957a). Information Theory and Statistical Mechanics I. *Physical Review* 106, pp. 62–30.

Jaynes, E.T. (1957b). Information Theory and Statistical Mechanics II. *Physical Review* 108, pp. 171–90.

Jaynes, E.T. (1968). Prior Probabilities. *IEEE Transactions on Systems Science and Cybernetics* SEC-4, pp. 227–41.

Jeffrey, Richard C. (1965). *The Logic of Decision.* 1st edition. McGraw-Hill Series in Probability and Statistics. New York: McGraw-Hill.

Jeffrey, Richard C. (1970). Dracula Meets Wolfman: Acceptance vs. Partial Belief. In: *Induction, Acceptance, and Rational Belief.* Ed. by M. Swain. Dordrecht: Reidel, pp. 157–85.

Jeffrey, Richard C. (1983). *The Logic of Decision.* 2nd edition. Chicago: University of Chicago Press.

Jeffrey, Richard C. (1987). Indefinite Probability Judgment: A Reply to Levi. *Philosophy of Science* 54, pp. 586–91.

Jeffrey, Richard C. (1993). Causality and the Logic of Decision. *Philosophical Topics* 21, pp. 139–51.

Jeffrey, Richard C. (2004). *Subjective Probability: The Real Thing.* Cambridge: Cambridge University Press.

Johnson, W.E. (1932). Probability: The Deductive and Inductive Problems. *Mind* 41, pp. 409–23.

Joyce, James M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science* 65, pp. 575–603.

Joyce, James M. (1999). *The Foundations of Causal Decision Theory.* Cambridge: Cambridge University Press.

Joyce, James M. (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives* 19, pp. 153–78.

Joyce, James M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In: *Degrees of Belief.* Ed. by Franz Huber and Christoph Schmidt-Petri. Vol. 342. Synthese Library. Springer, pp. 263–97.

Joyce, James M. (2010). A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives* 24, pp. 281–323.

Kahneman, Daniel and Amos Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* XLVII, pp. 263–91.

Kaplan, Mark (1996). *Decision Theory as Philosophy.* Cambridge: Cambridge University Press.

Kelley, Mikayla (ms). On Accuracy and Coherence with Infinite Opinion Sets. Unpublished manuscript.

Kemeny, John G. (1955). Fair Bets and Inductive Probabilities. *The Journal of Symbolic Logic* 20, pp. 263–73.

Kemeny, John G. and Paul Oppenheim (1952). Degree of Factual Support. *Philosophy of Science* 19, pp. 307–24.

Keynes, John Maynard (1921). *Treatise on Probability.* London: Macmillan and Co., Limited.

Keynes, John Maynard (1923). *A Tract on Monetary Reform.* Macmillan and Co., Limited.

Keynes, John Maynard (1937). The General Theory of Employment. *The Quarterly Journal of Economics* 51, pp. 209–23.

Kierland, Brian and Bradley Monton (2005). Minimizing Inaccuracy for Self-locating Beliefs. *Philosophy and Phenomenological Research* 70, pp. 384–95.

Kim, Jaegwon (1988). What Is "Naturalized Epistemology"? *Philosophical Perspectives* 2, pp. 381–405.

Kim, Namjoong (2009). Sleeping Beauty and Shifted Jeffrey Conditionalization. *Synthese* 168, pp. 295–312.

Knight, Frank Hyneman (1921). *Risk, Uncertainty and Profit.* Boston and New York: Houghton Mifflin.

Kolmogorov, A.N. (1933/1950). *Foundations of the Theory of Probability.* Translation edited by Nathan Morrison. New York: Chelsea Publishing Company.

Konek, Jason and Benjamin A. Levinstein (2019). The Foundations of Epistemic Decision Theory. *Mind* 128, pp. 69–107.

Kornblith, Hilary (1993). Epistemic Normativity. *Synthese* 94, pp. 357–76.

Kraft, Charles H., John W. Pratt, and A. Seidenberg (1959). Intuitive Probability on Finite Sets. *The Annals of Mathematical Statistics* 30, pp. 408–19.

Krantz, D.H. et al. (1971). *Foundations of Measurement, Vol 1: Additive and Polynomial Representations*. Cambridge: Academic Press.

Kuhn, Thomas S. (1957). *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought.* New York: MJF Books.

Kyburg Jr, Henry E. (1961). *Probability and the Logic of Rational Belief.* Middletown: Wesleyan University Press.

Kyburg Jr, Henry E. (1970). Conjunctivitis. In: *Induction, Acceptance, and Rational Belief.* Ed. by M. Swain. Boston: Reidel, pp. 55–82.

Laddaga, R. (1977). Lehrer and the Consensus Proposal. *Synthese* 36, pp. 473–7.

Lance, Mark Norris (1995). Subjective Probability and Acceptance. *Philosophical Studies* 77, pp. 147–79.

Lange, Alexandra (2019). Can Data Be Human? The Work of Giorgia Lupi. *The New Yorker.* Published May 25, 2019.

Lange, Marc (2000). Is Jeffrey Conditionalization Defective by Virtue of Being Non-commutative? Remarks on the Sameness of Sensory Experience. *Synthese* 123, pp. 393–403.

Laplace, Pierre-Simon (1814/1995). *Philosophical Essay on Probabilities.* Translated from the French by Andrew Dale. New York: Springer.

Lehman, R. Sherman (1955). On Confirmation and Rational Betting. *Journal of Symbolic Logic* 20, pp. 251–62.

Lehrer, K. and Carl Wagner (1983). Probability Amalgamation and the Independence Issue: A Reply to Laddaga. *Synthese* 55, pp. 339–46.

Leitgeb, Hannes and Richard Pettigrew (2010a). An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science* 77, pp. 201–35.

Leitgeb, Hannes and Richard Pettigrew (2010b). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science* 77, pp. 236–72.

Lele, Subhash R. (2004). Evidence Functions and the Optimality of the Law of Likelihood. In: *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations.* Chicago: University of Chicago Press, pp. 191–216.

Levi, Isaac (1974). On Indeterminate Probabilities. *The Journal of Philosophy* 71, pp. 391–418.

Levi, Isaac (1980). *The Enterprise of Knowledge.* Boston: The MIT Press.

Levi, Isaac (1987). The Demons of Decision. *The Monist* 70, pp. 193–211.

Lewis, C.I. (1946). *An Analysis of Knowledge and Valuation.* La Salle, IL: Open Court.

Lewis, David (1971). Immodest Inductive Methods. *Philosophy of Science* 38, pp. 54–63.

Lewis, David (1976). Probabilities of Conditionals and Conditional Probabilities. *The Philosophical Review* 85, pp. 297–315.

Lewis, David (1979). Atittudes *de dicto* and *de se*. *The Philosophical Review* 88, pp. 513–43.

Lewis, David (1980). A Subjectivist's Guide to Objective Chance. In: *Studies in Inductive Logic and Probability.* Ed. by Richard C. Jeffrey. Vol. 2. Berkeley: University of California Press, pp. 263–94.

Lewis, David (1981a). Causal Decision Theory. *Australasian Journal of Philosophy* 59, pp. 5–30.

Lewis, David (1981b). 'Why Ain'cha Rich?' *Noûs* 15, pp. 377–80.

Lewis, David (1994). Humean Supervenience Debugged. *Mind* 103, pp. 473–90.

Lewis, David (1996). Desire as Belief II. *Mind* 105, pp. 303–13.

Lewis, David (2001). Sleeping Beauty: Reply to Elga. *Analysis* 61, pp. 171–6.

Lichtenstein, S., B. Fischoff, and L. Phillips (1982). Calibration of Probabilities: The State of the Art to 1980. In: *Judgment under Uncertainty: Heuristics and Biases.* Ed. by Daniel Kahneman, P. Slovic, and Amos Tversky. Cambridge: Cambridge University Press, pp. 306–34.

Lindley, Dennis V. (1982). Scoring Rules and the Inevitability of Probability. *International Statistical Review* 50, pp. 1–26.

Lindley, Dennis V. (1985). *Making Decisions.* 2nd edition. London: Wiley.

Liu, Liping and Ronald R. Yager (2008). Classic Works of the Dempster-Shafer Theory of Belief Functions: An Introduction. In: *Classic Works of the Dempster-Shafer Theory of Belief Functions.* Ed. by Ronald R. Yager and Liping Liu. Vol. 219. Studies in Fuzziness and Soft Computing. Berlin: Springer, pp. 1–34.

Locke, John (1689/1975). *An Essay Concerning Human Understanding.* Ed. by Peter H. Nidditch. Oxford: Oxford University Press.

MacFarlane, John (2005). Making Sense of Relative Truth. *Proceedings of the Aristotelian Society* 105, pp. 321–39.

Maher, Patrick (1993). *Betting on Theories.* Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge: Cambridge University Press.

Maher, Patrick (1996). Subjective and Objective Confirmation. *Philosophy of Science* 63, pp. 149–74.

Maher, Patrick (2002). Joyce's Argument for Probabilism. *Philosophy of Science* 96, pp. 73–81.

Maher, Patrick (2010). Explication of Inductive Probability. *Journal of Philosophical Logic* 39, pp. 593–616.

Makinson, David C. (1965). The Paradox of the Preface. *Analysis* 25, pp. 205–7.

Makinson, David C. (2011). Conditional Probability in the Light of Qualitative Belief Change. *Journal of Philosophical Logic* 40, pp. 121–53.

Mayo, Deborah (2018). *Statistical Inference as Severe Testing: How to Get beyond the Statistics Wars.* Cambridge: Cambridge University Press.

Mazurkiewicz, Stefan (1932). Zur Axiomatik der Wahrscheinlichkeitsrechnung. *Comptes rendues des séances de la Société des Sciences et des Lettres de Varsovie* 25, pp. 1–4.

Meacham, Christopher J. G. (2008). Sleeping Beauty and the Dynamics of De Se Beliefs. *Philosophical Studies* 138, pp. 245–70.

Meacham, Christopher J. G. (2010a). Unravelling the Tangled Web: Continuity, Internalism, Uniqueness, and Self-locating Belief. In: *Oxford Studies in Epistemology.* Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 3. Oxford: Oxford University Press, pp. 86–125.

Meacham, Christopher J.G. (2010b). Two Mistakes Regarding the Principal Principle. *British Journal for the Philosophy of Science* 61, pp. 407–31.

Meacham, Christopher J. G. (2016). Ur-Priors, Conditionalization, and Ur-Prior Conditionalization. *Ergo* 3, pp. 444–92.

Meacham, Christopher J.G. and Jonathan Weisberg (2011). Representation Theorems and the Foundations of Decision Theory. *Australasian Journal of Philosophy* 89, pp. 641–63.

Mellor, D.H. (2013). Review of *Probability in the Philosophy of Religion. Analysis* 73, pp. 548–54.

Moore, David S., George P. McCabe, and Bruce A. Craig (2009). *Introduction to the Practice of Statistics.* 6th edition. New York: W.H. Freeman and Company.

Moore, G.E. (1939). Proof of an External World. *Proceedings of the British Academy* 25, pp. 273–300.

Moss, Sarah (2012). Updating as Communication. *Philosophy and Phenomenological Research* 85, pp. 225–48.

Moss, Sarah (2015). Credal Dilemmas. *Noûs* 49, pp. 665–83.

Moss, Sarah (2018). *Probabilistic Knowledge.* Oxford: Oxford University Press.

Murphy, Allan H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology* 12, pp. 595–600.

Murphy, Allan H. and Robert L. Winkler (1977). Reliability of Subjective Probability Forecasts of Precipitation and Temperature. *Journal of the Royal Statistical Society, Series C* 26, pp. 41–7.

Neal, Radford M. (2006). *Puzzles of Anthropic Reasoning Resolved Using Full Non-indexical Conditioning.* Tech. rep. 0607. Department of Statistics, University of Toronto.

Neyman, J. and Egon Pearson (1967). *Joint Statistical Papers.* Cambridge: Cambridge University Press.

Nicod, Jean (1930). *Foundations of Geometry and Induction.* Translated by Philip Wiener. New York: Harcourt, Brace and Company.

Nozick, Robert (1969). Newcomb's Problem and Two Principles of Choice. In: *Essays in Honor of Carl G. Hempel.* Synthese Library. Dordrecht: Reidel, pp. 114–15.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349. DOI: 10.1126/science.aac4716.

Papineau, David (2012). *Philosophical Devices: Proofs, Probabilities, Possibilities, and Sets.* Oxford: Oxford University Press.

Paris, J.B. (1994). *The Uncertain Reasoner's Companion.* Cambridge: Cambridge University Press.

Pascal, Blaise (1670/1910). *Pensées*. Translated by W.F. Trotter. London: Dent.

Pearson, K., A. Lee, and L. Bramley-Moore (1899). Genetic (Reproductive) Selection: Inheritance of Fertility in Man. *Philosophical Transactions of the Royal Society A* 73, pp. 534–39.

Pedersen, Arthur Paul and Gregory Wheeler (2014). Demystifying Dilation. *Erkenntnis* 79, pp. 1305–42.

Peirce, Charles Sanders (1910/1932). Notes on the Doctrine of Chances. In: *Collected Papers of Charles Sanders Peirce*. Ed. by Charles Hartshorne and Paul Weiss. Cambridge, MA: Harvard University Press, pp. 404–14.

Peterson, Martin (2009). *An Introduction to Decision Theory*. Cambridge Introductions to Philosophy. Cambridge: Cambridge University Press.

Pettigrew, Richard (2013a). A New Epistemic Utility Argument for the Principal Principle. *Episteme* 10, pp. 19–35.

Pettigrew, Richard (2013b). Epistemic Utility and Norms for Credences. *Philosophy Compass* 8, pp. 897–908.

Pettigrew, Richard (2014). Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research* 92(1), pp. 35–59.

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

Pettigrew, Richard (2021). On the Expected Utility Objection to the Dutch Book Argument for Probabilism. *Noûs* 55, pp. 23–38.

Pettigrew, Richard and Michael G. Titelbaum (2014). Deference Done Right. *Philosophers' Imprint* 14, pp. 1–19.

Pettigrew, Richard and Jonathan Weisberg, eds (2019). *The Open Handbook of Formal Epistemology*. Published open access online by PhilPapers. URL: https://philpapers.org/rec/PETTOH-2.

Piccione, Michele and Ariel Rubinstein (1997). On the Interpretation of Decision Problems with Imperfect Recall. *Games and Economic Behavior* 20, pp. 3–24.

Pollock, John L. (2001). Defeasible Reasoning with Variable Degrees of Justification. *Artificial Intelligence* 133, pp. 233–82.

Popper, Karl R. (1935/1959). *The Logic of Scientific Discovery*. London: Hutchinson & Co.

Popper, Karl R. (1938). A Set of Independent Axioms for Probability. *Mind* 47, pp. 275–9.

Popper, Karl R. (1954). Degree of Confirmation. *British Journal for the Philosophy of Science* 5, pp. 143–9.

Popper, Karl R. (1955). Two Autonomous Axiom Systems for the Calculus of Probabilities. *British Journal for the Philosophy of Science* 6, pp. 51–7.

Popper, Karl R. (1957). The Propensity Interpretation of the Calculus of Probability and the Quantum Theory. *The Colston Papers* 9. Ed. by S. Körner, pp. 65–70.

Predd, J. et al. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory* 55, pp. 4786–92.

Pryor, James (2004). What's Wrong with Moore's Argument? *Philosophical Issues* 14, pp. 349–78.

Quinn, Warren S. (1990). The Puzzle of the Self-Torturer. *Philosophical Studies* 59, pp. 79–90.

Ramsey, Frank P. (1929/1990). General Propositions and Causality. In: *Philosophical Papers*. Ed. by D.H. Mellor. Cambridge: Cambridge University Press, pp. 145–63.

Ramsey, Frank P. (1931). Truth and Probability. In: *The Foundations of Mathematics and other Logic Essays*. Ed. by R.B. Braithwaite. New York: Harcourt, Brace and Company, pp. 156–98.

Rees, Martin (2000). *Just Six Numbers: The Deep Forces that Shape the Universe*. New York: Basic Books.

Reichenbach, Hans (1935/1949). *The Theory of Probability.* English expanded version of the German original. Berkeley: University of California Press.

Reichenbach, Hans (1938). *Experience and Prediction.* Chicago: University of Chicago Press.

Reichenbach, Hans (1956). The Principle of Common Cause. In: *The Direction of Time.* Berkeley: University of California Press, pp. 157–60.

Reiss, Julian and Jan Sprenger (2017). Scientific Objectivity. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/.

Renyi, Alfred (1970). *Foundations of Probability.* San Francisco: Holden-Day.

Resnik, Michael D. (1987). *Choices: An Introduction to Decision Theory.* Minneapolis: University of Minnesota Press.

Roche, William (2014). Evidence of Evidence Is Evidence under Screening-off. *Episteme* 11, pp. 119–24.

Roeper, P. and H. Leblanc (1999). *Probability Theory and Probability Logic.* Toronto: University of Toronto Press.

Rosenkrantz, Roger (1981). *Foundations and Applications of Inductive Probability.* Atascadero, CA: Ridgeview Press.

Rosenthal, Jeffrey S. (2009). A Mathematical Analysis of the Sleeping Beauty Problem. *The Mathematical Intelligencer* 31, pp. 32–37.

Royall, Richard M. (1997). *Statistical Evidence: A Likelihood Paradigm.* New York: Chapman & Hall/CRC.

Salmon, Wesley (1966). *The Foundations of Scientific Inference.* Pittsburgh: University of Pittsburgh Press.

Salmon, Wesley (1975). Confirmation and Relevance. In: *Induction, Probability, and Confirmation.* Ed. by Grover Maxwell and Robert M. Jr. Anderson. Vol. VI. Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press, pp. 3–36.

Savage, Leonard J. (1954). *The Foundations of Statistics.* New York: Wiley.

Savage, Leonard J. (1967). Difficulties in the Theory of Personal Probability. *Philosophy of Science* 34, pp. 305–10.

Schervish, M.J., T. Seidenfeld, and J.B. Kadane (2004). Stopping to Reflect. *Journal of Philosophy* 101, pp. 315–22.

Schervish, M.J., T. Seidenfeld, and J.B. Kadane (2009). Proper Scoring Rules, Dominated Forecasts, and Coherence. *Decision Analysis* 6, pp. 202–21.

Schick, Frederic (1986). Dutch Bookies and Money Pumps. *The Journal of Philosophy* 83, pp. 112–19.

Schoenfield, Miriam (2017). The Accuracy and Rationality of Imprecise Credences. *Noûs* 51, pp. 667–85.

Schwarz, Wolfgang (2010). *Lewis on Updating and Self-location.* Blog post available at URL: https://www.umsu.de/wo/2010/563.

Schwarz, Wolfgang (2018). Subjunctive Conditional Probability. *Journal of Philosophical Logic* 47, pp. 47–66.

Scott, Dana (1964). Measurement Structures and Linear Inequalities. *Journal of Mathematical Psychology* 1, pp. 233–47.

Seelig, C. (1956). *Albert Einstein: A Documentary Biography.* London: Staples Press.

Seidenfeld, Teddy (1986). Entropy and Uncertainty. *Philosophy of Science* 53, pp. 467–91.

Seidenfeld, Teddy, M.J. Schervish, and J.B. Kadane (2017). Non-Conglomerability for Countably Additive Measures That Are Not κ-Additive. *The Review of Symbolic Logic* 10, pp. 284–300.

Seidenfeld, Teddy and Larry Wasserman (1993). Dilation for Sets of Probabilities. *The Annals of Statistics* 21, pp. 1139–54.

Selvin, Steve (1975). A Problem in Probability. *The American Statistician* 29. Published among the Letters to the Editor, p. 67.

Shafer, Glenn (1976). *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press.

Shafer, Glenn (1981). Constructive Probability. *Synthese* 48, pp. 1–60.

Shapiro, Amram, Louise Firth Campbell, and Rosalind Wright (2014). *The Book of Odds.* New York: Harper Collins.

Shimony, Abner (1955). Coherence and the Axioms of Confirmation. *Journal of Symbolic Logic* 20, pp. 1–28.

Shimony, Abner (1988). An Adamite Derivation of the Calculus of Probability. In: *Probability and Causality.* Ed. by J.H. Fetzer. Dordrecht: Reidel, pp. 151–61.

Shogenji, Tomoji (2003). A Condition for Transitivity in Probabilistic Support. *British Journal for the Philosophy of Science* 54, pp. 613–16.

Shogenji, Tomoji (2012). The Degree of Epistemic Justification and the Conjunction Fallacy. *Synthese* 184, pp. 29–48.

Shope, R.K. (1978). The Conditional Fallacy in Contemporary Philosophy. *Journal of Philosophy* 75, pp. 397–413.

Shortliffe, E. and B. Buchanan (1975). A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences* 23, pp. 351–79.

Simpson, E.H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Series B* 13, pp. 238–41.

Skipper, Mattias and Jens Christian Bjerring (2022). Bayesianism for Nonideal Agents. *Erkenntnis* 87, pp. 93–115.

Skyrms, Brian (1980a). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws.* New Haven CT: Yale University Press.

Skyrms, Brian (1980b). Higher Order Degrees of Belief. In: *Prospects for Pragmatism.* Ed. by D.H. Mellor. Cambridge: Cambridge University Press, pp. 109–37.

Skyrms, Brian (1983). Three Ways to Give a Probability Function a Memory. In: *Testing Scientific Theories.* Ed. by John Earman. Vol. 10. Minnesota Studies in the Philosophy of Science. Minneapolis: University of Minnesota Press, pp. 157–61.

Skyrms, Brian (1987a). Coherence. In: *Scientific Inquiry in Philosophical Perspective.* Ed. by N. Rescher. Pittsburgh: University of Pittsburgh Press, pp. 225–42.

Skyrms, Brian (1987b). Dynamic Coherence and Probability Kinematics. *Philosophy of Science* 54, pp. 1–20.

Skyrms, Brian (2000). *Choice & Chance: An Introduction to Inductive Logic.* 4th edition. Stamford, CT: Wadsworth.

Smith, C.A.B. (1961). Consistency in Statistical Inference and Decision. *Journal of the Royal Statistical Society, Series B* 23, pp. 1–25.

Smithies, Declan (2015). Ideal Rationality and Logical Omniscience. *Synthese* 192, pp. 2769–93.

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

Spohn, Wolfgang (2012). *The Laws of Belief: Ranking Theory and its Philosophical Applications.* Oxford: Oxford University Press.

Staffel, Julia (2019). *Unsettled Thoughts: Reasoning and Uncertainty in Epistemology.* Oxford: Oxford University Press.

Stalnaker, Robert C. (1972/1981). Letter to David Lewis. In: *Ifs: Conditionals, Belief, Decision, Chance, and Time.* Ed. by W. Harper, Robert C. Stalnaker, and G. Pearce. Dordrecht: Reidel, pp. 151–2.

Stalnaker, Robert C. (2008). *Our Knowledge of the Internal World.* Oxford: Oxford University Press.

Stalnaker, Robert C. (2011). Responses to Stoljar, Weatherson and Boghossian. *Philosophical Studies* 155, pp. 467–79.

Stefánsson, H. Orri (2017). What Is "Real" in Probabilism? *Australasian Journal of Philosophy* 95, pp. 573–87.

Stefánsson, H. Orri (2018). On the Ratio Challenge for Comparativism. *Australasian Journal of Philosophy* 96, pp. 380–90.

Stephenson, Todd A. (2000). *An Introduction to Bayesian Network Theory and Usage.* Tech. rep. 03. IDIAP.

Sturgeon, Scott (2010). Confidence and Coarse-grained Attitudes. In: *Oxford Studies in Epistemology.* Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 3. Oxford: Oxford University Press, pp. 126–49.

Suppes, Patrick (1955). The Role of Subjective Probability and Utility in Decision-making. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* pp. 61–73.

Suppes, Patrick (1969). *Studies in the Methodology and Foundations of Science.* Berlin: Springer.

Suppes, Patrick (1974). *Probabilistic Metaphysics.* Uppsala: University of Uppsala Press.

Suppes, Patrick (2002). *Representation and Invariance of Scientific Structures.* Stanford, CA: CSLI Publications.

Tal, Eyal and Juan Comesaña (2017). Is Evidence of Evidence Evidence? *Noûs* 51, pp. 95–112.

Talbott, William J. (1991). Two Principles of Bayesian Epistemology. *Philosophical Studies* 62, pp. 135–50.

Talbott, William J. (2005). Review of "Putting Logic in its Place: Formal Constraints on Rational Belief". *Notre Dame Philosophical Reviews.* URL: http://ndpr.nd.edu/.

Talbott, William J. (2016). Bayesian Epistemology. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/.

Teller, Paul (1973). Conditionalization and Observation. *Synthese* 26, pp. 218–58.

Tennant, Neil (2017). Logicism and Neologicism. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/win2017/entries/logicism/.

Tentori, Katya, Vincenzo Crupi, and Selena Russo (2013). On the Determinants of the Conjunction Fallacy: Probability versus Inductive Confirmation. *Journal of Experimental Psychology: General* 142, pp. 235–55.

Titelbaum, Michael G. (2008). The Relevance of Self-locating Beliefs. *Philosophical Review* 117, pp. 555–605.

Titelbaum, Michael G. (2010). Not Enough There There: Evidence, Reasons, and Language Independence. *Philosophical Perspectives* 24, pp. 477–528.

Titelbaum, Michael G. (2013a). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief.* Oxford: Oxford University Press.

Titelbaum, Michael G. (2013b). Ten Reasons to Care about the Sleeping Beauty Problem. *Philosophy Compass* 8, pp. 1003–17.

Titelbaum, Michael G. (2015a). Continuing On. *Canadian Journal of Philosophy* 45, pp. 670–91.

Titelbaum, Michael G. (2015b). Rationality's Fixed Point (Or: In Defense of Right Reason). In: *Oxford Studies in Epistemology.* Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 5. Oxford University Press, pp. 253–94.

Titelbaum, Michael G. (2016). Self-locating Credences. In: *The Oxford Handbook of Probability and Philosophy.* Ed. by Alan Hájek and Christopher R. Hitchcock. Oxford: Oxford University Press, pp. 666–80.

Titelbaum, Michael G. (2019). Precise Credences. In: *The Open Handbook of Formal Epistemology.* Ed. by Richard Pettigrew and Jonathan Weisberg. The PhilPapers Foundation, pp. 1–56.

Tversky, Amos and Daniel Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* 185, pp. 1124–31.

Tversky, Amos and Daniel Kahneman (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90, pp. 293–315.

Tversky, Amos and Daniel Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5, pp. 297–323.

Tversky, Amos and Derek J. Koehler (1994). Support Theory: A Nonextensional Representation of Subjective Probability. *Psychological Review* 101, pp. 547–67.

van Enk, Steven J. (2015). Betting, Risk, and the Law of Likelihood. *Ergo* 2, pp. 105–21.

van Fraassen, Bas C. (1980). Rational Belief and Probability Kinematics. *Philosophy of Science* 47, pp. 165–87.

van Fraassen, Bas C. (1981). A Problem for Relative Information Minimizers. *British Journal for the Philosophy of Science* 32, pp. 375–79.

van Fraassen, Bas C. (1982). Rational Belief and the Common Cause Principle. In: *What? Where? When? Why?* Ed. by Robert McLaughlin. Dordrecht: Reidel, pp. 193–209.

van Fraassen, Bas C. (1983). Calibration: A Frequency Justification for Personal Probability. In: *Physics Philosophy and Psychoanalysis.* Ed. by R. Cohen and L. Laudan. Dordrecht: Reidel, pp. 295–319.

van Fraassen, Bas C. (1984). Belief and the Will. *The Journal of Philosophy* 81, pp. 235–56.

van Fraassen, Bas C. (1989). *Laws and Symmetry.* Oxford: Clarendon Press.

van Fraassen, Bas C. (1995). Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies* 77, pp. 7–37.

van Fraassen, Bas C. (1999). Conditionalization: A New Argument For. *Topoi* 18, pp. 93–6.

van Fraassen, Bas C. (2005). Conditionalizing on Violated Bell's Inequalities. *Analysis* 65, pp. 27–32.

Van Horn, Kevin S. (2003). Constructing a Logic of Plausible Inference: A Guide to Cox's Theorem. *International Journal of Approximate Reasoning* 34, pp. 3–24.

Venn, John (1866). *The Logic of Chance.* London, Cambridge: Macmillan.

Vineberg, Susan (2011). Dutch Book Arguments. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Summer 2011. URL: https://plato.stanford.edu/archives/su2011/entries/dutch-book/.

von Mises, Richard (1928/1957). *Probability, Statistics and Truth.* (English edition of the original German *Wahrscheinlichkeit, Statistik und Wahrheit.*) New York: Dover.

von Neumann, J. and O. Morgenstern (1947). *Theory of Games and Economic Behavior.* 2nd edition. Princeton, NJ: Princeton University Press.

Vranas, Peter B.M. (2004). Hempel's Raven Paradox: A Lacuna in the Standard Bayesian Solution. *British Journal for the Philosophy of Science* 55, pp. 545–60.

Wainer, Howard (2011). *Uneducated Guesses: Using Evidence to Uncover Misguided Education Policies.* Princeton, NJ: Princeton University Press.

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities.* London: Chapman and Hall.

Wasserstein, Ronald L. and Nicole A. Lazar (2016). The ASA's Statement on p-values: Context, Process, and Purpose. *The American Statistician.* DOI: 10.1080/00031305.2016.1154108.

Weatherson, Brian (2002). Keynes, Uncertainty and Interest Rates. *Cambridge Journal of Economics* 26, pp. 47–62.

Weatherson, Brian (2011). Stalnaker on Sleeping Beauty. *Philosophical Studies* 155, pp. 445–56.

Weatherson, Brian (2015). For Bayesians, Rational Modesty Requires Imprecision. *Ergo* 2, pp. 529–45.

Weatherson, Brian and Andy Egan (2011). Epistemic Modals and Epistemic Modality. In: *Epistemic Modality.* Ed. by Andy Egan and Brian Weatherson. Oxford: Oxford University Press, pp. 1–18.

Weintraub, Ruth (2001). The Lottery: A Paradox Regained and Resolved. *Synthese* 129, pp. 439–49.

Weirich, Paul (2012). Causal Decision Theory. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2012. URL: https://plato.stanford.edu/archives/win2012/entries/decision-causal/.

Weisberg, Jonathan (2007). Conditionalization, Reflection, and Self-Knowledge. *Philosophical Studies* 135, pp. 179–97.

Weisberg, Jonathan (2009). Varieties of Bayesianism. In: *Handbook of the History of Logic.* Ed. by Dov. M Gabbya, Stephan Hartmann, and John Woods. Vol. 10: Inductive Logic. Oxford: Elsevier.

White, Roger (2005). Epistemic Permissiveness. *Philosophical Perspectives* 19, pp. 445–59.

White, Roger (2006). Problems for Dogmatism. *Philosophical Studies* 131, pp. 525–57.

White, Roger (2010). Evidential Symmetry and Mushy Credence. In: *Oxford Studies in Epistemology.* Ed. by Tamar Szabó Gendler and John Hawthorne. Vol. 3. Oxford: Oxford University Press, pp. 161–86.

Williams, J. Robert G. (ms). A Non-Pragmatic Dominance Argument for Conditionalization. Unpublished manuscript.

Williams, J. Robert G. (2016). Probability and Non-Classical Logic. In: *Oxford Handbook of Probability and Philosophy.* Ed. by Alan Hájek and Christopher R. Hitchcock. Oxford: Oxford University Press.

Williamson, Timothy (2000). *Knowledge and its Limits.* Oxford: Oxford University Press.

Williamson, Timothy (2007). How Probable Is an Infinite Sequence of Heads? *Analysis* 67, pp. 173–80.

Wittgenstein, Ludwig (1921/1961). *Tractatus Logico-Philosophicus.* Translated by D.F. Pears and B.F. McGuinness. London: Routledge.

Wong, S.M. et al. (1991). Axiomatization of Qualitative Belief Structure. *IEEE Transactions on Systems, Man and Cybernetics* 21, pp. 726–34.

Woolston, Chris (2015). Psychology Journal Bans *P* Values. *Nature* 519.9. URL: https://doi.org/10.1038/519009f.

Wroński, Leszek and Godziszewski, Michał Tomasz (2017). *The Stubborn Non-probabilist—'Negation Incoherence' and a New Way to Block the Dutch Book Argument.* International Workshop on Logic, Rationality and Interaction. Berlin, Heidelberg: Springer, pp. 256–67.

Yalcin, Seth (2012). A Counterexample to Modus Tollens. *Journal of Philosophical Logic* 41, pp. 1001–24.

Yule, G.U. (1903). Notes on the Theory of Association of Attributes in Statistics. *Biometrika* 2, pp. 121–34.

Zynda, Lyle (1995). Old Evidence and New Theories. *Philosophical Studies* 77, pp. 67–95.

Zynda, Lyle (2000). Representation Theorems and Realism about Degrees of Belief. *Philosophy of Science* 67, pp. 45–69.