Logical ignorance and logical learning

Richard Pettigrew*

May 9, 2020

Abstract

According to certain normative theories in epistemology, rationality requires us to be logically omniscient. Yet this prescription clashes with our ordinary judgments of rationality. How should we resolve this tension? In this paper, I focus particularly on the logical omniscience requirement in Bayesian epistemology. Building on a key insight by Ian Hacking (1967), I develop a version of Bayesianism that permits logical ignorance. This includes: (i) an account of the synchronic norms that govern a logically ignorant individual at any given time; (ii) an account of how we reduce our logical ignorance by learning logical facts and how we should update our credences in response to such evidence; and (iii) an account of when logical ignorance is irrational and when it isn't. At the end, I explain why the requirement of logical omniscience remains true of ideal agents with no computational, processing, or storage limitations.

According to certain normative theories in epistemology, rationality requires us to be logically omniscient. We should believe or be certain of every logical truth; we should believe something if and only if we believe everything logically equivalent to it, and we should be confident in something to exactly the same degree we're confident in anything logically equivalent to it; if we believe something, we should believe all its logical consequences, and we should be no more confident in something than we are in any of its logical consequences; and so on. For instance, if you think that we should believe everything we have reason to believe, and if you think that the truth of each tautology gives us reason to believe it,

*I am grateful to the following people for helping me to think through the issues in this paper: Branden Fitelson, Catrin Campbell-Moore, Jason Konek, Johannes Stern, Kevin Dorst, Julia Staffel, Jennifer Carr, Sophie Horowitz, David Christensen, and Kenny Easwaran. I started working on the paper after David Thorstad, Ned Hall, and Branden Fitelson invited me to speak at their workshop on bounded rationality at Harvard in October 2019, and I am very grateful to them for the prompt and to Branden for his comments on the paper. It had its next outing at the Eastern APA in January 2020, and I thank the audience there for very helpful comments. The referees from this journal also helped to improve the paper considerably.

then you think we should believe every tautology (Smithies, 2015). Or, if you are a Bayesian epistemologist, one of the central tenets of your creed is Probabilism, and that demands we assign maximal credence—namely, 1, or 100%—to all logical truths and minimal credence—namely, 0, or 0%—to all logical falsehoods; and it demands that we assign at most as much credence to something as to its logical consequences (Hacking, 1967; Jeffrey, 1992; Garber, 1983; Earman, 1992).[1] In this paper, I'll focus primarily on the logical omniscience that Bayesianism demands of our credences; but I'll briefly consider the case of belief in my conclusion.

Now, at least in the Bayesian case, we do not simply demand logical omniscience on a whim; we have our reasons. For one thing, in many standard applications of the Bayesian machinery, the assumption is natural. Suppose, for instance, we're trying to use it to decide between two hypotheses about the chance with which a coin will land heads when it is tossed. There's a fixed, easily surveyable set of hypotheses; there is a fixed, easily enumerable set of possible bodies of evidence we might obtain after observing the coin toss for, say, twenty tosses; and the logical relationships between the pieces of evidence, between the hypotheses, and between the evidence and the hypotheses are pretty transparent to someone who understands the setup. In this case, logical omniscience tells us that we must be certain that, if the coin didn't land heads on the first toss, it landed tails; it tells us to be at least as confident that it landed heads on the first toss as we are that it landed heads on the first two tosses; and so on. In these cases, logical omniscience seems a reasonable requirement. And it was such cases where Bayesianism first found application.

More generally, and indeed with greater normative force, we are led to demand logical omniscience by the best available arguments for credal norms, namely, the Dutch Book Argument, due to Frank P. Ramsey and Bruno de Finetti, and the Accuracy Dominance Argument, due to James Joyce (Ramsey, 1926 [1931]; de Finetti, 1937 [1980]; Joyce, 1998). The former claims that my credences are irrational if they're *incoherent* (or *Dutchbookable*), and they're incoherent (or Dutchbookable) if there's a series of bets, each of which my credences require me to accept, but which, when taken together, are guaranteed to lose me money. It then argues that, if I am not logically omniscient, I'm incoherent in this sense, and therefore irrational. The latter says that my credences are irrational if they're *accuracy dominated*, and they're accuracy dominated if there are alternative credences over the same propositions that are guaranteed to be strictly more accurate than mine. It then argues that, if I am not logically omniscient, I'm accuracy dominated in this sense, and therefore irrational.

[1] Probabilism says that (i) your credence in a logical truth should be 1 ($c(\top) = 1$), (ii) your credence in a logical falsehood should be 0 ($c(\bot) = 0$), and (iii) your credence in a disjunction should be the sum of your credences in the disjuncts less your credence in their conjunction ($c(A \vee B) = c(A) + c(B) - c(A \,\&\, B)$).

So the simplest uses of Bayesian epistemology, and the uses to which it was originally put in the philosophy of science, make the requirement of logical omniscience seem reasonable; and, what's more, our best arguments concerning credal norms justify it. However, our intuitive judgments about rationality are at odds with this prescription. Consider a first year logic student who has just learned the method of truth tables, and who is asked to determine the status of Peirce's Law, namely, $(((p \to q) \to p) \to p)$. They have learned enough logic to understand the truth functional properties of the material conditional, and they recognise what the well-formed formula just stated says. But they are surely not irrational if they are less than certain of that formula when they begin to draw up its truth table; and they are surely not irrational if they become steadily more and more certain of it as they secure a 'T' in the final column of each of the four rows.

Bayesianism is a well supported, successful, and pleasingly general theory of rational credence. How should we reconcile it with the verdicts of our intuitions about the rationality of logical uncertainty, which are widely held and resilient? Some philosophers respond to the tension by siding with logical omniscience, at least as a requirement of ideal rationality (Smithies, 2015). Others accept that Bayesianism has over played its hand here, and they try to weaken it in a way that preserves those features that have made it so successful (Hacking, 1967; Jeffrey, 1992; Garber, 1983; Gaifman, 2004). I'll follow the second route. I will adopt and expand Ian Hacking's approach to this issue (Hacking, 1967). Hacking applied his approach primarily to synchronic credal norms; I will expand it to diachronic norms as well; and I will explain how we might appeal to an application of Good's Value of Information Theorem (Good, 1967) in order to say when it is appropriate to criticize an individual for their logical ignorance, and to recover what is right about the demand for logical omniscience as a requirement of ideal rationality.

1 Hacking's insight

As noted above, the first obstacles for someone who wishes to reject the demand of logical omniscience are the Dutch Book Argument and the Accuracy Dominance Argument, each of which is usually taken to establish Probabilism and with it logical omniscience. In this section, we draw on an insight due originally to Ian Hacking (1967) to show that, when correctly understood, both arguments in fact establish a weaker norm, Personal Probabilism, which only entails standard Probabilism, and thus logical omniscience, in certain situations. This alternative norm matches our intuitive judgments about credal rationality much better.

Hacking originally explored his insight in the case of the Dutch Book Argument. Daniel Garber (1983) then elaborated on it and described some

of its consequences. Robbie Williams (2018) then applied the insight to the accuracy argument in a related but slightly different context from ours, namely, the case in which you are uncertain which logic governs the propositions to which you assign credences, but you are certain what follows from what according to the consequence relations of each logic you consider possible. Here, we will mainly be interested in the case in which you are certain that classical logic is correct, but you are not certain about what follows from what according to its consequence relation.

To understand Hacking's insight, note that the Dutch Book Argument and the Accuracy Dominance Argument are both based on what we might call *dominance reasoning*. In the Dutch Book Argument, we show that, if your credences violate Probabilism, they require you to make a dominated series of choices; that is, they require you to make each of a series of choices where there is an alternative series of choices you could have made instead that would leave you better off at all possible worlds: in particular, you must choose to accept each of a series of bets where refusing each bet would leave you better off. In the Accuracy Dominance Argument, we show that, if your credences violate Probabilism, they are accuracy dominated; that is, there is an alternative set of credences that is more accurate than yours at all possible worlds. Thus, in both arguments, we have to specify which possible worlds we mean to include among those at which the utility of the outcomes of the actions in the Dutch Book Argument and the accuracy of credences in the Accuracy Dominance Argument will be evaluated.

In what follows, we'll assume that classical logic is the correct logic and we'll say of a world that it is logically possible world if it is possible by the lights of classical propositional logic. Now, Hacking noted of the Dutch Book Argument that, if it is to establish Probabilism, it must assume that the possible worlds include *at most* those that are logically possible. And, if we wish to prove the Converse Dutch Book Theorem, which is intended to establish that no credence function that satisfies Probabilism is incoherent, then we need to include *at least* those logically possible worlds. Thus, in the Dutch Book Argument, we typically take them to be *exactly* the logically possible worlds. And the same is true of the Accuracy Dominance Argument. In both cases, we can secure a credal norm that does not demand logical omniscience if we permit more than just the logically possible worlds.

This was Hacking's insight. It immediately raises the question: Which worlds should be included? Which are the *dominance-relevant worlds*? That is, what is the smallest set of worlds, $\mathcal{W}$, for which dominance reasoning holds? That is, for which $\mathcal{W}$ is it the case that you are irrational if you choose one option when there is an alternative option you might have chosen that is better than your option at all worlds in $\mathcal{W}$? The Dutch Book Argument and the Accuracy Dominance Argument for Probabilism work only if $\mathcal{W}$ contains only logically possible worlds. Is that a reasonable as-

sumption? In the remainder of this section, I'll argue that it is not.

Let's begin by supposing I tell you that the only dominance-relevant world is the actual world. That is, I claim that you are irrational if you choose an option when there's an alternative that is, as a matter of actual fact, better than yours. This immediately strikes you as wrong. You think I must be mixing up the best thing to do with the rational thing to do. It is surely possible to be rational and yet choose an option that is, as a matter of actual fact, not the best. I take your point. So I expand the set of dominance-relevant worlds to include not only the actual world, but also all the metaphysically possible worlds as well. That is, I now claim you are irrational if you choose an option when there's an alternative that is, as a matter of metaphysical necessity, better than yours. This seems less problematic, but still wrong. After all, I don't know all the a posteriori but metaphysically necessary truths; that is, I don't know which are the metaphysically possible worlds; that is, I don't have sufficient evidence to pin them down; I don't have sufficient evidence to rule out all the metaphysically impossible worlds. So surely I can't be irrational simply because I choose something that is dominated relative to this set of worlds.

In both cases, our judgments about the account of the dominance-relevant worlds seem to be based on the following very rough account of rationality. To be rational is to do the best that you can within the bounds of your limited resources and as judged from your limited perspective. If you choose one option when there is another that is actually better, it is possible to see from an alternative, better informed perspective that you are not doing as well as you might; but if it isn't possible to see this from your limited perspective, you are not irrational. And similarly if you choose an option where there is another that is better at all metaphysically possible worlds. Again, from a better informed perspective, we can see that there is something that is guaranteed to be better than your choice; but if you can't see that from your more limited perspective, you aren't irrational for choosing as you do.

This suggests the notion of possibility that Hacking introduces, namely, *personal possibility*.[2] Roughly speaking, a world is personally possible for a particular individual at a particular time if by this time this individual hasn't ruled it out by their experiences, their logical reasoning, their conceptual thinking, their insights, their emotional reactions, or whatever other cognitive activities and processes can rule out worlds for an individual. For instance, if I learn from a visual experience that there's a goat in front of me, this rules out all worlds at which there isn't. If I learn by testimony that South Africa won the 2019 Rugby World Cup, then that rules out

[2]Robbie Williams (2018, 132) takes up the same idea under the name *doxastic possibility*. He leaves it as a primitive, undefined concept. Hacking himself comes to identify personal possibilities with epistemic possibilities (Hacking, 1967, Section 6). We'll have more to say about this below.

all worlds at which England won it, providing I've already learned that no more than one team can win and that South Africa is a different team from England, and thereby already ruled out worlds at which England won too. If I learn by philosophical reasoning and reflection that the Law of the Excluded Middle is true, and I know that Y is the disjunction of X and the negation of X, that rules out all worlds at which Y is false. If I learn an instance of Peirce's Law by the method of truth tables, that rules out all worlds at which that instance is false. And so on.

Pace Williams and Hacking, we should distinguish personally possible worlds from doxastically or epistemically possible ones. A world is doxastically possible if it makes true everything you believe, while it is epistemically possible if it makes true everything you know. In both cases, we need to posit categorical doxastic attitudes, such as belief and knowledge, to define the notion. Yet one might be sceptical about these. And in any case our approach does not depend on their existence. A world is personally possible not if it makes true all you know or believe, but rather if it hasn't been ruled out by your cognitive activities and processes.[3]

According to Hacking (and to Williams and to me), the dominance-relevant worlds are precisely those that are personally possible for the individual in question at the time in question. If you choose option a, and yet b is better at all metaphysically possible worlds, you aren't necessarily irrational, for you might not have enough evidence to rule out certain metaphysically impossible worlds—if you don't know Stokely Carmichael was Kwame Ture, then you might not have ruled out a world at which Carmichael was a protégé of Ella Baker, but Ture wasn't. And, similarly, if b is better at all logically possible worlds, you aren't necessarily irrational, for you might not have reasoned hard enough or long enough to rule out certain logically impossible worlds—you might never have ruled out a world in which the axioms of arithmetic are true, but there are just finitely many primes. However, if b is better at all personally possible worlds, then it is irrational to choose a. After all, a is worse not only from some external, better informed point of view; it's also better from your own limited perspective. And that's what it means to be irrational.

I conclude this section by considering two questions that arise. First: What exactly are these logically impossible worlds that I wish to include among our personal possibilities? Like Graham Priest (1997), it seems to me that we might extend any of the standard metaphysical accounts of logically possible worlds to the realm of the logically impossible: for instance, Yagisawa (1988) extends Lewis' modal realism in this way, positing concrete logically impossible worlds; Jago (2012) constructs ersatz impossible worlds out of positive and negative facts; it's easy to see how to extend

[3]These personally possible worlds share much in common with Graham Priest's *open worlds* (Priest, 2005).

the linguistic ersatzism view that possible worlds are sentences in natural language or interpreted formal languages; and so on. For my purposes here, I remain agnostic about the metaphysics of these worlds. I do this not because I have no view on the matter, but because the account I give does not depend on which view we adopt—it will work just as well for any account with a handful of minimal features, which I will now enumerate. First, I need that, for any personal possibility and anything to which I might assign a credence, the possibility determines whether the object of the credence is true or false. Now, just as there are many accounts of the metaphysics of personal possibilities, so there are many accounts of the objects of credence. Chalmers (2011) enumerates a few. He rejects Russellian structured propositions and sets of metaphysically possible worlds because of a credal version of Frege's puzzle. He rejects sentences in natural language, because he takes it to be possible to have contentful thoughts while lacking linguistic ability. He considers sentences in the language of thought, Fregean thoughts, and pairs consisting of Russellian propositions together with a guise under which that proposition is apprehended in thought. He favours sets of centred worlds. As for the metaphysics of personal possibilities, I remain agnostic about the metaphysics of the objects of our credences.[4] All that I need is that the objects of credence are the sorts of things made true or false by the personal possibilities. More precisely, in order to get our framework up and running, I need:

(i) for each individual and each time, a set $\mathcal{W}$ of worlds that are the personal possibilities for that individual at that time;

(ii) for each individual and each time, a set $\mathcal{F}$ of credal objects to which that individual assigns credences at that time;

(iii) for each X in $\mathcal{F}$ and w in $\mathcal{W}$, X is true or false at w;

(iv) it is not necessary that, for every subset $\mathcal{S} \subseteq \mathcal{W}$, there is X in $\mathcal{F}$ that is true at all and only the worlds in $\mathcal{S}$.

Any metaphysical account of personal possibilities and credal objects that satisfies (i)-(iv) can serve our purpose.

The second question I wish to consider before we move on: How do we rule out particular personally possible worlds? I gave some examples above: we have perceptual experiences and emotional reactions and insights; we undertake logical reasoning and conceptual thinking. Each of these is a cognitive process or activity. When does an individual rule out a particular personally possible world by using such a process or activity? I won't have much to say about this here. Rather, I'll take the standard approach of the Bayesian, who is usually not so interested in what Jonathan

[4]However, it is worth noting that Chalmers' centred worlds won't work for us, since they are all logically if not metaphysically possible.

Weisberg (2009) calls the *input problem*, namely, how the agent acquires the new information she does, and focusses instead on the norms that govern how they should respond to that information. I'll assume only that these inputs, whatever they are and however they are delivered to you, serve to eliminate personally possible worlds. That is, after these learning episodes, your set of personally possible worlds shrinks. Thus, on this picture, while the inputs that give rise to logical and empirical learning might be different—following a deductive inference or using a test for satisfiability in one case, observing the world around you or dredging up a memory of a past observation in the other—the effect they have on you is of the same sort. In each case, learning helps you whittle down the personal possibilities, allowing you to home in on the actual world.

You might fill out the details of this account in a number of different ways. For instance, you might say that a cognitive process terminates in a set of personally possible worlds—a logical deduction terminates in the set of personally possible worlds at which its conclusion is true; a perceptual experience terminates in the set of personally possible worlds at which you have that perceptual experience; and so on. Then you might say that a cognitive process is reliable if it tends to result in sets of personally possible worlds that contain the actual world. And finally, you might say that I have ruled out a personally possible world by using a cognitive process if (i) that process is reliable and (ii) the world in question isn't in the set of personally possible worlds that the process produces. But this is just one way. What I say in the remainder of the paper does not rely on this.

2 The Dutch Book Argument

So the Dutch Book and the Accuracy Dominance Arguments for Probabilism both rely on dominance reasoning. This means that, in order to draw specific conclusions from them, we need to specify $\mathcal{W}$, the dominance-relevant set of worlds. Here, I'll walk through the structure of both arguments, but I'll leave $\mathcal{W}$ as a variable we can specify later. As we'll see, we get different conclusions for different specifications of $\mathcal{W}$.

The Dutch Book Argument begins with a claim about which bets your credences require you to accept.[5]

> **Favourable Bets** If you have credence p in X, you are required to pay any amount less than $\$pS$ for a bet that pays $\$S$ if X is true and $\$0$ if X is false.

[5]Dutch Book arguments have attracted many objections over the years. For presentations of the arguments and the objections, see (Hájek, 2008; Vineberg, 2016; Pettigrew, 2020). I present arguments of this sort here not because I endorse them but because they are one of the main sources of pressure to accept logical omniscience.

Such a bet is called a *\$S-bet on X*, and \$S is the *stake*. Thus, Favourable Bets says that, if you are 70% confident that it will rain tomorrow, then your credence requires you to pay \$5 or \$6.49 or any amount less than \$7 for a \$10-bet on rain tomorrow. In this example, the stake of the bet is \$10.

Now suppose your credence function is c, which is defined on $\mathcal{F}$, the set of credal objects you entertain. Then we say *c is incoherent over $\mathcal{W}$* if there is a series of bets on the credal objects in $\mathcal{F}$ such that (i) the credences that c assigns require you to accept each of those bets individually and (ii) taken together, those bets lose you money at every world in $\mathcal{W}$. Thus, if c is incoherent over $\mathcal{W}$, then c requires you to make a dominated series of choices. It requires you to accept each bet in the series, but refusing each of them would leave you better off. Thus, c requires you to do something irrational. So, if we count credences as irrational that require you to make an irrational series of choices, then credence functions that are incoherent over $\mathcal{W}$ are irrational.

It just remains to identify which credence functions are incoherent over $\mathcal{W}$. To do this, we turn to a theorem due to de Finetti (1974). To state it, we need some new terminology.

- Suppose w is a possible world in $\mathcal{W}$. Then define the function $v_w : \mathcal{F} \to \{0, 1\}$ as follows:

$$v_w(X) = \begin{cases} 1 & \text{if } X \text{ is true at } w \\ 0 & \text{if } X \text{ is false at } w \end{cases}$$

  We call v_w the *valuation function for w on $\mathcal{F}$*.

- Let $\mathcal{W}_\mathcal{F}$ be the set of valuation functions for the worlds in $\mathcal{W}$ defined on $\mathcal{F}$. That is,

$$\mathcal{W}_\mathcal{F} = \{v_w : \mathcal{F} \to \{0, 1\} \mid w \in \mathcal{W}\}$$

- Let $\mathcal{W}_\mathcal{F}^+$ be the convex hull of $\mathcal{W}_\mathcal{F}$. That is,

  (i) $\mathcal{W}_\mathcal{F}^+$ is convex, so that whenever c and c' are in $\mathcal{W}_\mathcal{F}^+$, so is $\lambda c + (1 - \lambda)c'$ for any $0 \leq \lambda \leq 1$,

  (ii) $\mathcal{W}_\mathcal{F}$ is a subset of $\mathcal{W}_\mathcal{F}^+$, and

  (iii) $\mathcal{W}_\mathcal{F}^+$ is the smallest set for which (i) and (ii) hold.

Now we can state the theorem:

Theorem 1 (de Finetti) *A credence function c on $\mathcal{F}$ is incoherent over $\mathcal{W}$ iff c is not in $\mathcal{W}_\mathcal{F}^+$.*

Now, if $\mathcal{W}$ is just the classically possible worlds, then a further theorem due to de Finetti gives us Probabilism:

Theorem 2 (de Finetti) *If $\mathcal{W}$ is the set of classically possible worlds, then c is in $\mathcal{W}^+_{\mathcal{F}}$ iff c satisfies Probabilism.*

Thus, if the dominance-relevant worlds are exactly the logically possible worlds, then the incoherent credence functions are exactly those that violate Probabilism, and we can derive that norm. But if they extend beyond that, we cannot. Instead, in the general case, we derive the norm of Personal Probabilism.

> **Personal Probabilism** If c is your credence function and $\mathcal{W}$ is the set of personally possible worlds for you, then it ought to be that c is in $\mathcal{W}^+_{\mathcal{F}}$.

We'll explore some of the consequences of this norm in particular cases below.

2.1 The Accuracy Dominance Argument

While the Dutch Book Argument judges a non-probabilistic credence function irrational because it requires you to make a dominated series of choices, the Accuracy Dominance Argument judges it irrational because it is itself dominated relative to a particular measure of epistemic goodness, namely, accuracy. That is, there is an alternative credence function that is more accurate at all worlds.

At the heart of the argument are the measures of inaccuracy. These take a credence function and a possible world and they measure how inaccurate that credence function is at that world. There are a number of different ways of specifying which measures of inaccuracy are legitimate. For the sake of concreteness, let's focus on the so-called *additive continuous strictly proper inaccuracy measures*, though what I say will apply to other classes as well (Joyce, 2009; Predd et al., 2009; Pettigrew, 2016). I'll briefly define these here, but the details aren't essential to the rest of the discussion. A *scoring rule* is a function $\mathfrak{s}$ that takes a truth value, represented by 1 or 0, and a credence p and returns a measure, $\mathfrak{s}(1, p)$ or $\mathfrak{s}(0, p)$, of the inaccuracy of having credence p in a credal object with that truth value. A scoring rule is *continuous* if $\mathfrak{s}(1, x)$ and $\mathfrak{s}(0, x)$ are both continuous functions of x. A scoring rule is *strictly proper* if each credence expects itself to have lowest inaccuracy by the lights of that scoring rule: that is, if $p\mathfrak{s}(1, x) + (1 - p)\mathfrak{s}(0, x)$ is minimized uniquely, as a function of x, at $x = p$. An inaccuracy measure $\mathfrak{I}$ is an *additive continuous strictly proper inaccuracy measure* if there is a continuous strictly proper scoring $\mathfrak{s}$, and the inaccuracy that $\mathfrak{I}$ assigns to a credence function c at a world w is the sum of the scores given by $\mathfrak{s}$ at w to the credences that c assigns: that is, $\mathfrak{I}(c, w) = \sum_{X \in \mathcal{F}} \mathfrak{s}(v_w(X), c(X))$, where v_w is the valuation function for w that we defined above. Thus, a central premise in the version of the Accuracy Dominance Argument we're spelling out here is this:

Propriety Any legitimate measure of inaccuracy is an additive continuous strictly proper inaccuracy measure.

Now, according to the Accuracy Dominance Argument, a credence function c on $\mathcal{F}$ is irrational if it is accuracy dominated over $\mathcal{W}$ relative to all legitimate inaccuracy measures, and it is accuracy dominated over $\mathcal{W}$ relative to an inaccuracy measure $\mathfrak{I}$ if there is an alternative credence function $c^\star$ that is less inaccurate at every world in $\mathcal{W}$, when inaccuracy is measured by $\mathfrak{I}$. That is, c is irrational if there is $c^\star$, also defined on $\mathcal{F}$, such that $\mathfrak{I}(c^\star, w) < \mathfrak{I}(c, w)$ for all w in $\mathcal{W}$.

It just remains to identify which credence functions are accuracy dominated over $\mathcal{W}$. To do this, we turn to a theorem due to Predd et al. (2009) that generalizes another theorem due to de Finetti (1974).

Theorem 3 (Predd, et al.) *Suppose $\mathfrak{I}$ is an additive continuous strictly proper inaccuracy measure. Then a credence function c on $\mathcal{F}$ is accuracy dominated over $\mathcal{W}$ relative to $\mathfrak{I}$ iff c is not in $\mathcal{W}^+_{\mathcal{F}}$.*

Thus, again, if $\mathcal{W}$ is the set of logically possible worlds, then we can appeal to Theorem 2 to derive Probabilism; and if not we can still derive Personal Probabilism.

2.2 Personal Probabilism

What does Personal Probabilism demand? This depends, of course, on the set of worlds that are personally possible for you. Let's illustrate with an example. Above, we briefly saw David Chalmers arguing that, if we take sentences in the subject's language to be the objects of credence, we fail to give a fully general theory, since we can ascribe credences even to subjects with no linguistic ability. However, for those with linguistic ability, sentences in their language are a decent proxy for whatever we do end up taking the objects of credence to be—they can be paired reasonably straightforwardly with the true objects of credence. In this example, then, we'll assume that the objects of your credences are indeed paired up with sentences. In fact, throughout, we'll suppose you have credences only in three sentences, A, B, and C. And we suppose that C is in fact the disjunction of A and B: that is, C is $A \vee B$.

At first, we assume you don't know that C is $A \vee B$. Indeed, you know nothing of the logical connections between the three sentences A, B, and C. This might be because you haven't yet attended to their logical forms. As a result, for you, there are eight personally possible worlds, namely, the eight different ways to assign truth values to A, B, and C (shown on the left of the table), and these correspond to eight different valuation functions (shown on the right).

$\mathcal{W}$	A	B	$C\ (= A \vee B)$	$\mathcal{W}_{\mathcal{F}}$	A	B	$C\ (= A \vee B)$
w_1	T	T	T	v_{w_1}	1	1	1
w_2	T	T	F	v_{w_2}	1	1	0
w_3	T	F	T	v_{w_3}	1	0	1
w_4	T	F	F	v_{w_4}	1	0	0
w_5	F	T	T	v_{w_5}	0	1	1
w_6	F	T	F	v_{w_6}	0	1	0
w_7	F	F	T	v_{w_7}	0	0	1
w_8	F	F	F	v_{w_8}	0	0	0

It turns out that, when $\mathcal{W} = \{w_1, \ldots, w_8\}$ is your set of personally possible worlds, no credence function is incoherent or accuracy-dominated; all are rationally permitted. That is because each possible assignment of credences is some convex combination of the valuation functions in $\mathcal{W}_{\mathcal{F}}$; so each credence function is in $\mathcal{W}_{\mathcal{F}}^{+}$, and therefore, by Theorems 1 and 3, it is neither incoherent nor accuracy dominated. So, in particular, a credence function that assigns lower credence to the disjunction $A \vee B$ than to either of its disjuncts A and B is not incoherent and its not accuracy dominated.

Next, let's suppose you learn that C is the disjunction of A and B. But how? After all, you don't assign any credence to a sentence, $C = (A \vee B)$, that states that identity. But not all learning must happen like that. Learning the truth of a sentence that you already entertain is just one way of learning. In general, you learn when you rule out particular worlds that were previously personally possible for you; that is, learning restricts the set of worlds that are personally possible for you. In particular, in this case, what you learned rules out worlds w_2, w_4, w_6, and w_7, and leaves worlds w_1, w_3, w_5, and w_8 as still personally possible for you. So after this learning episode, your set of personally possible worlds is $\mathcal{W}' = \{w_1, w_3, w_5, w_8\}$. Now, it's straightforward to show that the credence functions that satisfy Personal Probabilism when $\mathcal{W}'$ is the set of your personally possible worlds are exactly those c such that $c(A), c(B) \leq c(A \vee B) \leq c(A) + c(B)$.

Now, suppose that, later still, you learn that A and B are incompatible. That is, what you learn rules out world w_1, leaving w_3, w_5, and w_8 as your personally possible worlds. So $\mathcal{W}''_{\mathcal{F}} = \{w_3, w_5, w_8\}$. Then the credence functions that are neither incoherent nor accuracy dominated are now those that satisfy $c(A) + c(B) = c(A \vee B)$.

As Hacking notes, in general, we can recover something of standard Probabilism from Personal Probabilism (Hacking, 1967, Section 10):

- if you learn that $\top$ is a logical truth—that is, if it is true in all personally possible worlds—then Personal Probabilism requires $c(\top) = 1$, which is also what Probabilism requires;

- if you learn that $\bot$ is a logical falsehood—so that it is false at all personally possible worlds—then Personal Probabilism requires $c(\bot) =$

0, which is also what Probabilism requires;

- if you learn that $A \vee B$ is the disjunction of A and B, and $A \,\&\, B$ is their conjunction, then Personal Probabilism requires

$$c(A \vee B) = c(A) + c(B) - c(A \,\&\, B)$$

which is also what Probabilism requires.

Sinan Dogramaci (2018) worries about norms like Personal Probabilism. If Logicality is the traditional Bayesian requirement to be certain of any logical truth, let Personal Logicality be the requirement to be certain of any credal object that is true at all of your personally possible worlds. This is a consequence of Personal Probabilism. But Dogramaci worries that it's vacuous: "any proposition, as far as I can see, could be such that it might be false for you or me or whoever, and we have to say there is nothing irrational about it" (Dogramaci, 2018, 118). I think this commits a quantifier shift fallacy. Dogramaci is right that there is no credal object $\top$, such that, for all individuals, they are rationally required to be certain of $\top$. But that is not sufficient to show that Personal Logicality is vacuous. That would require that, for any individual and any credal object X, they are not required to be certain of X. But Dogramaci hasn't shown that. And indeed, it's false on the view I'm developing. Consider the following credal object: $(p \to q) \to (\neg q \to \neg p)$. I have taught logic many times; I have derived this formula in a number of different proof systems and I have verified via the method of truth tables that it is guaranteed to be true. I have, therefore, ruled out all personally possible worlds at which it is false. As a result, Personal Logicality demands that I am certain that $(p \to q) \to (\neg q \to \neg p)$. Nonetheless, it is possible for me to violate that norm. Suppose, for instance, that normally the cognitive process of deductive reasoning, which terminates in a set of personally possible worlds, also causes me to become certain of any credal object that is true at all of those worlds. However, sometimes, that causal pathway is interrupted; sometimes it malfunctions and gives me a credence of 0.5 instead. In such a case, I would violate Personal Logicality. In general, the point is that, while the process by which you rule out personally possible worlds will have close causal connections with the process by which you set your credences, they are nonetheless separate processes and thus can come apart. And when they do, that creates the possibility that you might assign credence less than 1 to a credal object that is true at all personally possible worlds that you haven't ruled out.

3 Logical learning

In the previous section, we rehearsed the synchronic or static credal norms that we obtain if we run the Dutch Book Argument and Accuracy Dominance Argument using personally possible worlds rather than logically

possible ones: we justify Personal Probabilism. We saw how learning more and thereby ruling out more personally possible worlds gave rise to increasingly strong synchronic norms. But we didn't discuss how we should update from the prior credences we have before we learn to our new posterior credences after we learn. We turn to that question in this section.

3.1 The two ways to learn

Now, there are two ways a learning episode can narrow down your set of personally possible worlds. The first is the most standard, and it is the one that Garber (1983) considers in his treatment of the problem of old evidence. In these episodes, we learn something with certainty that we have previously considered and to which we already assign a credence. That is, we learn of some credal object in $\mathcal{F}$ that it is true. For instance, I might have an intermediate credence in a particular instance of Peirce's Law, and then I might come to learn it with certainty on the basis of my logic tutor's testimony. Or we might learn a logical relationship, such as that the hypothesis H entails the evidence E, where I have a credence in the proposition $H \models E$ that I have thereby learned. In these cases, the diachronic story is simple: we have good pragmatic and epistemic arguments for updating by conditionalizing. David Lewis shows that, if you plan to update in any way other than by conditionalizing, there is a Dutch strategy against you—that is, a set of bets you'll accept before you learn anything, and another set of bets that you'll accept whatever you learn, and together these bets lose you money at all dominance-relevant possible worlds (Lewis, 1999). Peter M. Brown shows that, if you know you'll make a decision after receiving some evidence and you're planning how you should respond to that evidence so that you'll currently expect your future decision to be best, then you should update by conditionalization (Brown, 1976). Hilary Greaves and David Wallace show that, if you are planning how to update in response to new evidence you'll get, you expect conditionalization to be the most accurate method by which to do it (Greaves & Wallace, 2006). And Ray Briggs and I show that, if you assess the accuracy of your prior and your planned posteriors together, summing their individual accuracies, the only pairs that aren't accuracy dominated are those where your planned posteriors are the result of conditionalizing your prior on your evidence (Briggs & Pettigrew, 2018).

On the second way of learning, the learning episode directly eliminates personally possible worlds. It doesn't do so by teaching you that some credal object you entertain is true. Rather, it teaches you directly that some personally possible worlds are not actual. This is the sort of learning that took place in the move from $\mathcal{W}$ to $\mathcal{W}'$ and from $\mathcal{W}'$ to $\mathcal{W}''$ above. This sort of learning is much less often considered. One exception is the so-called 'superconditioning' treatment of the sort of empirical learning experience

for which Richard Jeffrey formulated his Probability Kinematics.[6] Jeffrey was interested in cases in which you have a particular perceptual experience but there is no proposition that you entertain and to which you previously assigned a credence that completely captures what you learn through that experience. Jeffrey suggests that, in such a case, you do not learn a proposition; rather, your evidence places some constraints on certain of your posterior credences. Probability Kinematics is then the updating rule you should follow to determine how to update the remaining credences to make them cohere with the ones fixed by the evidential constraints. Van Fraassen showed that you can similarly consider this as a case in which there is a proposition that you learn, but it's not one to which you initially assigned a credence: under certain conditions, including van Fraassen's own Reflection Principle, this approach and Jeffrey's are equivalent. Thus, on this view, the sort of empirical learning that Jeffrey wishes to treat with his Probability Kinematics is analogous to the sort of logical learning that we consider: both rule out possible worlds, but not necessarily by teaching you a proposition you previously entertained.

How should we respond to such learning? Neither Brown's pragmatic argument, nor Greaves and Wallace's accuracy argument will help us here, since they require you to assess the expected value of something from the point of view of your prior credences. But, in the case in question, the prior isn't defined on all of the personal possibilities that are required to define that expected value. For instance, to calculate an updating plan you might execute in the move from $\mathcal{W}$ to $\mathcal{W}'$, you'd need to assign prior credences to $w_1, \ldots, w_8$. But you don't. You assign credences only to A, which corresponds to the set $\{w_1, w_2, w_3, w_4\}$ of personally possible worlds, B, which corresponds to $\{w_1, w_2, w_5, w_6\}$, and C, which corresponds to $\{w_1, w_3, w_5, w_7\}$. However, Lewis' pragmatic argument and the accuracy argument that Briggs and I offer do apply. We'll see the updating norms they entail below.

3.2 Why we need both ways

Before that, however, I'd like to explain why both sorts of learning are essential to a comprehensive account of logical learning. It is because neither can account for all the important cases on their own. First, take Garber's central case, where you learn a fact about logical entailment, such as $H \models E$. Garber wishes to claim that, in many cases, old evidence can

[6]Jeffrey formulates Probability Kinematics in (Jeffrey, 1965). Bas van Fraassen (1980) and Brian Skyrms (1987) both show that Jeffrey's rule can be recovered as a form of superconditioning, providing you obey the Reflection Principle, as well as a condition called Sufficiency; Diaconis & Zabell (1982, Theorem 2.1) provide necessary and sufficient conditions for superconditioning. See (Konek, 2019) and (Wilson, ms) for more recent treatments of Jeffrey's Probability Kinematics and its relationship to superconditioning.

nonetheless support a new hypothesis if what the scientist learns is not the old evidence, but rather the logical relationship between the hypothesis and the evidence. Suppose our scientist has only credences in H and E—in Garber's example, following Glymour (1980, 85-6), H is Einstein's gravitational field equations, and E is the anomalous advance of the perihelion of Mercury. So, $\mathcal{F} = \{H, E\}$ and the personally possible worlds are given in the following table:

	H	E
w_1	T	T
w_2	T	F
w_3	F	T
w_4	F	F

Now, suppose we try to model learning $H \models E$ as eliminating world w_2 at which H is true and E false. Then that will not capture the full strength of $H \models E$, since a logical consequence claim is a modal claim. It does not reduce to the material implication $H \rightarrow E$, but learning that eliminates precisely w_2. What's more, this representation will not serve Garber's needs. Suppose that, like Einstein himself, I'm certain of E, because it is old evidence. Then my credence in w_2 is already 0. Thus, eliminating personally possible world w_2 will not lead me to update my credences at all, and in particular won't lead me to raise my credence in H. On the other hand, if I include $H \models E$ in $\mathcal{F}$, then it's quite possible that, by learning that logical fact, I can change my credence in H. For now the personally possible worlds are given in the following table:

	H	E	$H \models E$
w_1	T	T	T
w_2	T	T	F
w_3	T	F	T
w_4	T	F	F
w_5	F	T	T
w_6	F	T	F
w_7	F	F	T
w_8	F	F	F

Even if we follow Garber and eliminate w_3 because it is incompatible with the meaning of the consequence relation, and even if we eliminate w_4, w_7, and w_8 as well because E is false at those worlds, we are still left w_1, w_2, w_5, and w_6 as personally possible worlds. And there are priors that are mixtures of their valuation functions and such that conditionalizing on $H \models E$ raises your credence in H, as Garber requires.

Next, consider someone who has credences in H, E, and $H \models E$, but doesn't yet realise what $H \models E$ means. That is, she doesn't know how

it relates to its component parts, H and E (just as, in our example above, you didn't realise how $A \vee B$ related to A and to B). Then there is nothing she entertains now that she might learn with certainty that we can count as her learning the meaning of $H \models E$. To do that, she must eliminate w_3 and nothing more. But there is no credal object in her set $\mathcal{F} = \{H, E, H \models E\}$ that is false and w_3 and only at w_3. Of course, you might retort that $\overline{H \,\&\, \overline{E} \,\&\, H \models E}$ is false at w_3 and only at w_3. And while that is true, if we were to introduce that credal object into $\mathcal{F}$, the set of personally possible worlds would expand, and you'd first have to learn what $\overline{H \,\&\, \overline{E} \,\&\, H \models E}$ says before you can know that learning it with certainty will rule out exactly world w_3. And so the problem arises again: there is nothing you entertain that is true only at personally possible worlds at which $\overline{H \,\&\, \overline{E} \,\&\, H \models E}$ is related to its component parts in the requisite way. The upshot is that we need to be able to have learning experiences that rule out personally possible worlds directly, rather than by teaching you with certainty something you currently entertain, and thereby ruling out all personally possible worlds at which it is false.

3.3 How to update after you learn

In the previous section, we modelled learning as a process of whittling down your set of personally possible worlds, and we noted that this might happen when you learn the truth of something that you already have an opinion about; but it might also happen when your learning directly rules out personally possible worlds even though you couldn't articulate what you'd learned using those credal objects to which you assign credences. In this section, we ask how we should respond when we learn in these different ways.

Bayesian norms of updating are often stated as if they govern only actual updating behaviour. That is, we talk as if the norm tells you how you should update on the evidence you actually receive. However, the best arguments for those norms instead govern not only how you actually update on the evidence you actually receive, but also how you would update were you to receive different evidence. That is, these arguments establish norms that govern not only your actual updating behaviour, but your updating plans or intentions or dispositions. To specify your updating plan at a particular time, we first specify a partition $\mathcal{E} = \{E_1, \ldots, E_k\}$ of the set of your personally possible worlds at this time; this includes the pieces of evidence

you might obtain by some specific later time.[7] Thus, if you've just asked a yes/no question, $\mathcal{E}$ might partition the worlds into those where you receive a positive answer, those where you hear a negative, and those where the respondent tells you they don't know. Or, if you're about to check the time on a digital 24-hour clock showing hours and minutes, $\mathcal{E}$ will be the 1,440-cell partition of the worlds into those in which the clock shows '00:00', those in which it shows '00:01', and so on. An updating rule c' is then a function that takes each E_i in $\mathcal{E}$ and returns a credence function c'_i, which is the posterior that the rule endorses as a response to E_i.

In the Dutch Book argument for Personal Probabilism, we show that a credence function c is irrational by showing that c is *synchronically incoherent*: that is, there is a book B of bets, each of which c requires you to accept, such that the bets in B taken together will lose money at all personally possible worlds. In the Dutch Book argument for an updating norm, we show that a prior credence function c together with an updating rule c' on a partition $\mathcal{E} = \{E_1, \ldots, E_k\}$ is irrational by showing that (c, c') is *diachronically incoherent*: that is,

(i) there is a book B of bets, each of which c requires you to accept, and

(ii) for each E_i in $\mathcal{E}$, there is a book B_i of bets, each of which c'_i requires you to accept,

such that

(a) for any E_i in $\mathcal{E}$ and any personally possible world w in E_i, the bets in B and in B_i taken together lose you money at w.

Now, the question is: which prior-rule pairings are diachronically incoherent? In the standard diachronic Dutch Book argument for Bayesian conditionalization, this is usually asked when, for each E_i in $\mathcal{E}$, there is a X_i in $\mathcal{F}$ such that X_i is true at precisely the personally possible worlds in E_i. In this case, we say that X_i *represents* E_i; and when each E_i in $\mathcal{E}$ is represented by a credal object in $\mathcal{F}$, we say that $\mathcal{E}$ *is represented in* $\mathcal{F}$. To state our characterization of the prior-rule pairings that are diachronically incoherent when $\mathcal{E}$ is represented in $\mathcal{F}$, we must say what it means for a rule c' to be a conditionalizing rule for c.

Definition 1 *Suppose that, for each* $1 \leq i \leq k$, X_i *in* $\mathcal{F}$ *represents* E_i. c' is a conditionalizing rule for c *if, for each* E_i *in* $\mathcal{E}$, *if* $c(X_i) > 0$, *then* $c'_i(-) = c(-|X_i)$.

[7] We say that $\mathcal{E} = \{E_1, \ldots, E_k\}$ is a partition of $\mathcal{W}$ iff

(i) each E_i is a subset of $\mathcal{W}$;

(ii) together, they cover all of $\mathcal{W}$—that is, $E_1 \cup \ldots \cup E_k = \mathcal{W}$;

(iii) they are disjoint—that is, $E_i \cap E_j = \emptyset$, for any $1 \leq i \neq j \leq k$.

Then we have:

Theorem 4 *Suppose $\mathcal{E}$ is represented in $\mathcal{F}$. Then (c, c') is diachronically incoherent over $\mathcal{W}$ iff c' is not a conditionalizing rule for c.*

But we are also interested in cases in which, for some or even all E_i in $\mathcal{E}$, there is no X_i in $\mathcal{F}$ such that X_i is true at precisely the worlds in E_i. That is, $\mathcal{E}$ is not represented in $\mathcal{F}$. To say which prior-rule pairings are diachronically incoherent in this case, we must say what it means for a rule c' to a be a superconditioning rule for c.

Definition 2 c' is a superconditioning rule for c *if, for each w in $\mathcal{W}$, there is $0 \leq \lambda_w \leq 1$ such that $\sum_{w \in \mathcal{W}} \lambda_w = 1$ and*

(i) $c(-) = \sum_{w \in \mathcal{W}} \lambda_w v_w(-)$, *and*

(ii) *for E_i in $\mathcal{E}$, if $\sum_{w \in E_i} \lambda_w > 0$,*

$$c'_i(-) = \frac{\sum_{w \in E_i} \lambda_w v_w(-)}{\sum_{w \in E_i} \lambda_w}$$

Theorem 5 *(c, c') is diachronically incoherent over $\mathcal{W}$ iff c' is not a superconditioning rule for c.*

This furnishes us with an argument for the following norm:

> **Superconditionalization** If c is your prior and c' is your updating plan, then c' should be a superconditioning rule for c.

To understand superconditioning rules, it's best to see one in action. Suppose, as above, that $\mathcal{F} = \{A, B, A \vee B\}$. And suppose that you know that $A \vee B$ is the disjunction of A and B. Thus, your set of personally possible worlds is $\mathcal{W}'_{\mathcal{F}} = \{w_1, w_3, w_5, w_8\}$. Now, let's suppose that you're constructing a truth table that will teach you either $E_1 = \{w_1\}$, which says that A and B are both true together, or $E_2 = \{w_3, w_5, w_8\}$, which says that they are not both true together. Which updating rule c' should you adopt for responding this new evidence? As we saw above, at the prior time, Personal Probabilism demands only that $c(A), c(B) \leq c(A \vee B) \leq c(A) + c(B)$. At the posterior time, if you learn E_1, then that becomes your set of personally possible worlds, and Personal Probabilism demands that your posterior assigns $c'_1(A) = c'_1(B) = c'_1(A \vee B) = 1$; on the other hand, if you learn E_2, then that becomes your set of personally possible worlds, and Personal Probabilism demands that $c'_2(A) + c'_2(B) = c'_2(A \vee B)$. Now consider the following prior-rule pair (c, c'); the prior satisfies the synchronic constraints imposed by Personal Probabilism at the earlier time, and the two possible posteriors both satisfy the synchronic constraints imposed at the later time.

	A	B	$A \vee B$
c	$\frac{1}{4}$	$\frac{2}{3}$	$\frac{3}{4}$
c'_1	1	1	1
c'_2	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{7}{10}$

Then we can see that c' is a superconditioning rule for c by setting the weights as follows:

λ_{w_1}	λ_{w_3}	λ_{w_5}	λ_{w_8}
$\frac{2}{12}$	$\frac{1}{12}$	$\frac{6}{12}$	$\frac{3}{12}$

Then

(i) $c(-) = \frac{2}{12}v_{w_1} + \frac{1}{12}v_{w_3} + \frac{6}{12}v_{w_5} + \frac{3}{12}v_{w_8}$

(ii) $c'_1(-) = 1v_{w_1} + 0v_{w_3} + 0v_{w_5} + 0v_{w_8}$
$c'_2(-) = 0v_{w_1} + \frac{1}{10}v_{w_3} + \frac{6}{10}v_{w_5} + \frac{3}{10}v_{w_8}$

as required.

Having seen superconditioning in action, let's consider a diachronic version of the Accuracy Dominance Argument, which also establishes that norm. In the Accuracy Dominance Argument for Personal Probabilism, we show that a credence function c is irrational by showing that c is accuracy dominated: that is, there is an alternative $c^\star$ that is more accurate than c at all personally possible worlds. In the Accuracy Dominance Argument for an updating norm, we show that a prior c together with a rule c' defined on a partition $\mathcal{E}$ of your personally possible worlds is irrational by showing that the pair (c, c') is accuracy dominated: that is, there is an alternative pair $(c^\star, c^{\star\prime})$ such that, for each E_i in $\mathcal{E}$ and world w in E_i, the sum of the accuracy of $c^\star$ at w and the accuracy of $c_i^{\star\prime}$ at w exceeds the sum of the accuracy of c at w and the accuracy of c'_i at w. That is, if $\mathfrak{I}$ is an additive continuous strictly proper inaccuracy measure, then the inaccuracy of the pair (c, c') at world w in E_i is $\mathfrak{I}(c, w) + \mathfrak{I}(c'_i, w)$. So, (c, c') is accuracy dominated relative to $\mathfrak{I}$ iff there is $(c^\star, c^{\star\prime})$ such that, for all E_i in $\mathcal{E}$ and w in E_i,

$$\mathfrak{I}(c^\star, w) + \mathfrak{I}(c_i^{\star\prime}, w) < \mathfrak{I}(c, w) + \mathfrak{I}(c'_i, w)$$

Now, the question is: which prior-rule pairings are accuracy dominated relative to this way of measuring inaccuracy? Again, we have the case in which $\mathcal{E}$ is represented in $\mathcal{F}$:

Theorem 6 *Suppose $\mathcal{E}$ is represented in $\mathcal{F}$ and $\mathfrak{I}$ is a strictly proper inaccuracy measure. Then (c, c') is accuracy dominated over $\mathcal{W}$ relative to $\mathfrak{I}$ iff c' is not a conditionalizing rule for c.*

And we have the general case, where we don't assume that:

Theorem 7 *Suppose $\mathfrak{I}$ is a strictly proper inaccuracy measure. Then (c, c') is accuracy dominated over $\mathcal{W}$ relative to $\mathfrak{I}$ iff c' is not a superconditioning rule for c.*

This completes our extension of Hacking's project. Following his insight that the dominance-relevant worlds should include not only the logically possible worlds, but also the personally possible ones, we've shown the effect of this observation on the synchronic Dutch Book and Accuracy Dominance Arguments for Probabilism, and now we've seen how to update your credences in order to avoid diachronic incoherence and diachronic accuracy dominance. In the remainder of the paper, we turn to objections to this approach.

4 Objections

We'll consider two objections to the approach we've followed in this paper.

4.1 The threat of revenge

A common objection to the move from logically possible worlds to personally possible worlds is that it only buys the Bayesian a little time. After all, as we've seen, anyone with a credence function on $\mathcal{F}$ who satisfies our new personal version of Probabilism will have a credence function that is a mixture of the personally possible valuation functions over $\mathcal{F}$. So, if $\top$ is in $\mathcal{F}$ and $\top$ is true at all your personally possible worlds, Personal Probabilism requires you to be certain of $\top$. This is just what I called Personal Logicality when discussing Dogramaci's objection above.

Now, as we note when we introduce the problem of logical omniscience, all tautologies, however complex, are true at all logically possible worlds. However, for any particular tautology, and especially for complex ones, you might nonetheless not realise that it is a tautology and true at all logically possible worlds. And, if that's the case, rationality doesn't require you to be certain of it. Similarly, then, surely $\top$ might be true at all your personally possible worlds without you realising this. And, if that's the case, surely rationality doesn't require you to be certain of $\top$. The problem of logical omniscience, which says that it is too demanding to require you to be certain of all tautologies, returns in the guise of the problem of personal omniscience, which says that it is too demanding to require you to be certain of each credal object that is true at all your personally possible worlds. But the parallel is spurious. A credal object is true at all personally possible worlds only if your cognitive activities and processes have ruled out all worlds at which it is false. This cannot happen without you realising it. And if it's happened, then it is not too demanding to require you to be certain of that credal object.

4.2 The rationality of logical ignorance and logical sloth

You might worry that the solution to the problem of logical omniscience that we've been developing here renders logical sloth and extreme logical ignorance rationally permissible. That is, you might think that our framework furnishes us with no way to rationally criticise someone who fails to perform even basic logical reasoning to discover logical connections between the credal objects they entertain. On an account on which logical omniscience is rationally required, there is an incentive to learn more logical truths: while you will almost certainly remain ignorant of some logical truths, and therefore irrational on this account, you will at least become less and less irrational the more logical truths you learn—you will approach the ideal of rationality more closely. But on the Hacking-inspired account I have been developing, you might worry that there is no such incentive, and no way to criticise a person who does no logical reasoning.

However, as Hacking (1967) already observed in his original treatment, I. J. Good's Value of Information theorem already answers this objection (Good, 1967).[8] Here's the set-up for Good's theorem. Suppose you will face a decision between a range of options on Wednesday. And suppose that, on Wednesday, you'll pick an option that maximises your expected utility from the point of view of the credences you assign on that day. Now, suppose that, on Tuesday, you can choose whether or not to receive some information, and update on it by conditionalizing. Let's say that the information tells you which element of the partition $\mathcal{E} = \{E_1, \ldots, E_k\}$ contains the actual world. For ease of initial exposition, we'll assume that your credence function is defined on each personally possible world in $\mathcal{W}$—that is, for each personally possible world, there is a credal object in $\mathcal{F}$ that is true only at that world. Later, we'll explain how to lift that requirement. Good's theorem then says that, on Monday, you will expect yourself to choose a better outcome on Wednesday if you received the information on Tuesday and updated on that than if you didn't and retained your old credences. In symbols, if we say that a^c is an option that maximises expected utility relative to credence function c, and $a^{c'_i}$ is an option that maximises expected utility relative to c'_i, then

- by the lights of your prior c, the expected utility of not receiving the information is
$$\sum_{w \in \mathcal{W}} c(w) u(a^c, w)$$
- by the lights of your prior c, the expected utility of receiving the in-

[8] It's pretty clear that Savage already knew what's come to be known as Good's theorem when he was writing *The Foundations of Statistics* (Savage, 1954, Section 7.3).

formation and updating by conditionalizing on it is

$$\sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} c(w) u(a^{c'_i}, w)$$

Good's theorem says

Theorem 8 *If c is your prior, $\mathcal{E} = \{E_1, \ldots, E_k\}$ is a partition of $\mathcal{W}$ that is represented in $\mathcal{F}$, and c' is a conditionalizing rule for c, then*

$$\sum_{w \in \mathcal{W}} c(w) u(a^c, w) \leq \sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} c(w) u(a^{c'_i}, w)$$

with the inequality strict if there is E_i in $\mathcal{E}$ such that (i) $c(X_i) > 0$ *and* (ii) *the options that maximise expected utility by the lights of c'_i are different from those that maximise expected utility by the lights of c.*

This is often glossed: if the information on Tuesday is free, you're permitted to take it; if the information on Tuesday is free, and it might change your mind about the decision on Wednesday, then you're obliged to take it. Now, logical information isn't free. It takes time and cognitive resources to create or follow logical reasoning to obtain that information. So it might seem that Good's Theorem is not relevant here. But of course, it's not only if the information is free that you are obliged to take it. You're also obliged if its cost is lower than the expected difference between the utility of choosing after learning the information and the utility of choosing without learning the information. Thus, suppose it costs r utiles to learn which element of $\mathcal{E}$ contains the actual world. Then, you are obliged to take the information if

$$r < \sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} c(w) u(a^{c'_i}, w) - \sum_{w \in \mathcal{W}} c(w) u(a^c, w)$$

So there is a way to criticise someone who doesn't reason logically. If the gain in the expected utility of your future decisions exceeds the cost of that reasoning, then you're obliged to do it, and you can be rationally criticized if you don't.

What's more, there's an accuracy-based version of Good's Theorem.[9] Not only does learning increase the expected utility of your future decision making, it also increases the expected utility of your future credences. Suppose $\mathfrak{I}$ is a strictly proper inaccuracy measure. Then

- by the lights of your prior c, the expected inaccuracy of not receiving the information is

$$\sum_{w \in \mathcal{W}} c(w) \mathfrak{I}(c, w)$$

[9]This follows from the same reasoning used by Greaves & Wallace (2006) in their expected accuracy argument for conditionalization.

- by the lights of your prior c, the expected inaccuracy of receiving the information and updating by conditionalizing on it is

$$\sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} c(w)\mathfrak{I}(c'_i, w)$$

The accuracy version of Good's theorem says

Theorem 9 *Suppose $\mathfrak{I}$ is an additive continuous strictly proper inaccuracy measure. Then, if c is your prior, $\mathcal{E} = \{E_1, \ldots, E_k\}$ is a partition of $\mathcal{W}$ that is represented in $\mathcal{F}$, and c' is a conditionalizing rule for c, then*

$$\sum_{w \in \mathcal{W}} c(w)\mathfrak{I}(c, w) \geq \sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} c(w)\mathfrak{I}(c'_i, w)$$

with the inequality strict if there is E_i in $\mathcal{E}$ such that $c(X_i) > 0$ and $c \neq c'_i$.

Thus, again, if the information is free, you should take it. In order to say when you're obliged to pay a particular amount for the information, we need to specify an exchange rate between practical utility and epistemic utility. How much are you prepared to pay for a particular increase in your expected accuracy? But once we specify this, we can again rationally criticize someone who doesn't pay that amount.

So far, our presentation of the pragmatic and epistemic versions of Good's theorem have applied only in cases in which your prior is defined on each possible world in $\mathcal{W}$. But, as we have emphasised throughout, that is not the typical case. How might those arguments run if we do not assume that? If a credence function c satisfies Personal Probabilism, then it is a weighted average of the valuation functions of the individual's personally possible worlds. That is, there is a sequence $(\lambda_w)_{w \in \mathcal{W}}$ of weights, where $0 \leq \lambda_w \leq 1$ and $\sum_{w \in \mathcal{W}} \lambda_w = 1$ and

$$c(-) = \sum_{w \in \mathcal{W}} \lambda_w v_w(-)$$

We might think of λ_w as a possible credence you might assign to w if c is your credence function. It is a way of assigning credences to the personally possible worlds in $\mathcal{W}$ so that they are consistent with the credences that c already assigns. Now, note that the sequence $(\lambda_w)_{w \in \mathcal{W}}$ need not be unique. There might be an alternative sequence $(\delta_w)_{w \in \mathcal{W}}$ such that $0 \leq \delta_w \leq 1$ and $\sum_{w \in \mathcal{W}} \delta_w = 1$ and

$$c(-) = \sum_{w \in \mathcal{W}} \delta_w v_w(-) = \sum_{w \in \mathcal{W}} \lambda_w v_w(-)$$

Now, given any sequence $(\lambda_w)_{w \in \mathcal{W}}$ of weights, we can define expectations relative to them. For instance, the implicit expected utility of not receiving the evidence from partition $\mathcal{E} = \{E_1, \ldots, E_k\}$ relative to the weights $(\lambda_w)_{w \in \mathcal{W}}$ is

$$\sum_{w \in \mathcal{W}} \lambda_w u(a^c, w)$$

While the implicit expected utility of receiving the evidence relative to those weights is

$$\sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} \lambda_w u(a^{c'_i}, w)$$

And similarly for the implicit expected inaccuracy of both options. Now, again, we can show that

$$\sum_{w \in \mathcal{W}} \lambda_w u(a^c, w) \leq \sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} \lambda_w u(a^{c'_i}, w)$$

And that gives the general pragmatic version of Good's theorem. Similarly, we can show that

$$\sum_{w \in \mathcal{W}} \lambda_w \mathfrak{I}(c, w) \geq \sum_{E_i \in \mathcal{E}} \sum_{w \in E_i} \lambda_w \mathfrak{I}(c'_i, w)$$

which gives the general accuracy-based version of Good's theorem.

Given that, for many credence functions, there will be multiple sequences of weights, and therefore multiple assessments of expected utility relative to weights, and perhaps some that disagree on the ordering of the two options, we must say when an option is permitted and when it is mandated by such a credence function. I hope it is uncontroversial to say that an option is mandated by a credence function if it maximises expected utility relative to *all* possible sequences of weights. And this provides us with a way of rationally criticizing any individual who fails to seek out logical truths when doing so would increase their implicit expected utility or accuracy more than it would cost them to do so relative to all sequences of possible weights.

There are at least two compelling features of this approach to the problem of logical sloth. The first is that it matches our intuitive judgments about when an individual's logical ignorance is irrational. For instance, we judge a logically competent individual irrational if they are less than certain of $\neg(A \vee B) \rightarrow \neg A$. The reason is that the cost of reasoning to this conclusion will be extremely low. And, in general, we're more inclined to judge a logically ignorant person irrational if the logical truths of which they are less than certain are simpler and less costly to reason towards.

Notice that we thereby avoid the rather delicate business of determining when a logical contradiction is *subtle* or *complex* or *advanced*—in which case believing it or having high credence in it might be rationally permissible—and when it is *blatant* or *trivial* or *obvious*—in which case rationality might require us to disbelief it or have very low credence in it. And it manages to avoid it without making evaluations of rationality vague or indeterminate at the borderline between subtle and blatant contradictions (Lewis, 2004; Jago, 2014; Berto, 2014).

Our approach also allows us to make judgments of rationality that are appropriately relativised to an individual. For instance, it might allow us

to say that I am irrational if I'm less than certain of de Morgan's laws, because I've studied classical logic extensively, I've taught it, and I've worked through the derivation a hundred times; but a student in my first-year logic class, who has just learned the method of truth tables might not be irrational if they are less that certain of that law, because working through a derivation of it would cost that student much more than it would cost me.

This feature of our account allows us to answer an objection to Hacking's view raised by Jens Christian Bjerring and Mattias Skipper. They claim that, while Hacking successfully accounts for the rationality of logical ignorance, his picture fails to capture "what ordinary humans can and cannot infer given their limited cognitive resources" (Bjerring & Skipper, ta, 8). The diachronic norms for logical updating laid out in the previous section partially answer this by saying how to update when you do learn; and the norms for logical inquiry laid out in this section complete the answer by saying when you should put in the effort to learn. What's more, by appealing to Good's theorem, we secure a pleasingly general account. We needn't claim that all logical learning comes from following the rules of some deductive system, nor that there is some precise length of deduction within that system that we are capable of generating, nor that there is some impoverished language within which all of our logical reasoning at a given time must take place (Gaifman, 2004; Bjerring & Skipper, ta). Rather, we simply say that, however you might learn logical truths, you should choose to learn them iff doing so maximises your expected utility.

Nonetheless, the account I have given here does leave open a significant question, and it is perhaps the central question of Skipper and Bjerring's paper. How should we model actual logical reasoning epistemically? For me and for Hacking, this is part of what I called the input problem above, following Jonathan Weisberg. As I said there, I will not address it here. But perhaps I can adopt Skipper and Bjerring's answer. Thus, in the end, I think my project and theirs are complementary, not incompatible. They will tell you how to get to the point of learning a logical truth; I will then say how to update your credences on the basis of that achievement.

The second compelling feature of our approach is that it explains why individuals with none of the computational, processing, or storage limitations that we have should be logically omniscient, but not empirically omniscient. For such creatures, there is no cost whatsoever to undertaking the logical reasoning required to be omniscient. Providing logical knowledge can be acquired immediately and without incurring any opportunity or resource costs, Good's theorem shows that such an agent is rationally required to acquire all such knowledge. But for such creatures there is still a cost to gathering empirical evidence, and so they might not be rationally required to do that.

This provides a principled motivation for a norm of logical omniscience for ideal agents, but not a norm of empirical omniscience. It thereby pro-

vides an alternative justification for Declan Smithies' Asymmetry Thesis, at least for that sort of agent.

> **Asymmetry Thesis** Rationality requires omniscience and infallibility in logical domains but not empirical domains. (Smithies, 2015, 2772)

Smithies justifies that thesis in quite a different way, and by appealing to a reasons-based account of rationality. It is interesting that we can motivate the same principle from our teleological view of rationality.

5 Conclusion

The two major arguments for synchronic norms in credal epistemology, the Dutch Book and Accuracy Dominance Arguments—seem to demand logical omniscience. Yet that seems far too strong. Ian Hacking saw where the problem lay: since both rely on dominance reasoning, these two arguments require us to specify the set of dominance-relevant worlds; in the standard versions of the arguments, these are the logically possible worlds; in the case of the Dutch Book Argument, Hacking saw that they should instead be the personally possible worlds; and the same reasoning carries over to the Accuracy Dominance Argument. Replacing the logically possible worlds with the personally possible ones, we obtain a less demanding, more plausible norm, namely, Personal Probabilism. But this is a synchronic norm, so it does not tell us how to update when we learn logical facts and narrow our set of personally possible worlds. To do that, we introduced the diachronic versions of the Dutch Book and Accuracy Dominance Arguments and showed that they demand superconditioning as the rational updating rule. We then considered two objections to the framework that thereby grows out of Hacking's insight.

That framework contains multitudes. Beyond what we have seen here, Robbie Williams (2018) has applied Hacking's insight to the case in which the individual does not know which logic governs the objects to which she assigns credences, though she knows which follow from which according to the different logics she thinks are possible. But of course, an individual might be ignorant of both. Consider, for instance, an individual who doesn't know whether it is classical logic or the logic of paradox that governs her credal objects; and, moreover, she doesn't know all of the logical relationships that hold between those objects according to the two logics. Hacking's framework of personally possible worlds can represent such an individual and provide a norm that governs her credences at each time, and a norm that governs the way she updates on her evidence between those times.

Looking beyond credal epistemology, we can apply this framework to solve the problem of logical omniscience in the case of full belief as well. Building on work by Kenny Easwaran (2016) and Kevin Dorst (2017), Daniel Rothschild (2019) has provided notions of incoherence and accuracy dominance for sets of full beliefs. That is, (i) he specifies the bets that a set of beliefs require you to accept, and then says that a set is incoherent, and therefore irrational, if there is a series of bets each of which the set requires you to accept, but which taken together will lose you money at all possible worlds; and (ii) he specifies inaccuracy measures for sets of beliefs, and says that one is accuracy dominated relative to such a measure, and therefore irrational, if there is an alternative that is less inaccurate at all possible worlds. What's more, he has shown which sets of full beliefs are incoherent and which are accuracy dominated. If we take the dominance-relevant worlds to be the logically possible worlds, then a set of full beliefs is not incoherent just in case it is *almost Lockean complete* for some threshold $\frac{1}{2} < t < 1$—that is, there is a credence function c that satisfies Probabilism such that (i) if $c(X) > t$, then X is in the set of full beliefs and (ii) if $c(X) < t$, then X is not in the set of full beliefs. And exactly the same sets of beliefs are not accuracy dominated, again taking the logically possible worlds to be the dominance-relevant ones. Now, since c satisfies Probabilism, $c(\top) = 1 > t$ for any logical truth $\top$, and so full beliefs are required to be logically omniscient. However, if we appeal to Hacking's insight and let the dominance-relevant worlds instead be the personally possible ones as we did above, Rothschild's theorems establish that a set of full beliefs is not incoherent just in case there is a credence function c that satisfies Personal Probabilism such that (i) if $c(X) > t$, then X is in the set of full beliefs and (ii) if $c(X) < t$, then X is not in the set of full beliefs. Thus, we obtain synchronic norms for sets of full beliefs that do not demand logical omniscience. It is an interesting and open question which updating rules for sets of full beliefs are not vulnerable to diachronic versions of the Dutch Book and Accuracy Dominance arguments for full beliefs.

6 Proofs

Suppose:

- $\mathcal{F} = \{X_1, \ldots, X_n\}$
- $\mathcal{E} = \{E_1, \ldots, E_k\}$

Given a prior credence function c defined on $\mathcal{F}$, represent it by the vector

$$\langle c(X_1), \ldots, c(X_n) \rangle$$

Given a posterior credence function c'_i defined on $\mathcal{F}$, represent it by the vector

$$\langle c'_i(X_1), \ldots, c'_i(X_n) \rangle$$

Given a pair (c, c'), where c is a prior and c' is an updating rule on $\mathcal{E} = \{E_1, \ldots, E_k\}$, represent it by the vector

$$(c, c') = c \frown c'_1 \frown \ldots \frown c'_k$$

where $\frown$ is the concatenation operator between vectors, so that

$$\langle a_1, \ldots, a_n \rangle \frown \langle b_1, \ldots, b_n \rangle = \langle a_1, \ldots, a_n, b_1, \ldots, b_n \rangle$$

And thus

$$(c, c') = \langle c(X_1), \ldots, c(X_n), c'_1(X_1), \ldots, c'_1(X_n), \ldots, c'_k(X_1), \ldots, c'_k(X_n) \rangle$$

Given a pair (c, c') and a world w in E_i, let

$$(c, c')_w = v_w \frown c'_1 \frown \ldots \frown c'_{i-1} \frown v_w \frown c'_{i+1} \frown \ldots \frown c'_k$$

The key fact is this:

Lemma 10 *c' is a superconditioning rule for c iff c is in the convex hull of $\{(c, c')_w : w \in \mathcal{W}\}$.*[10]

Proof. First, left-to-right. Suppose c' is superconditioning for c. Then for each w in $\mathcal{W}$, there is $0 \leq \lambda_w \leq 1$ such that

(i) $c(-) = \sum_{w \in \mathcal{W}} \lambda_w v_w(-)$, and

(ii) for E_i in $\mathcal{E}$, if $\sum_{w \in E_i} \lambda_w > 0$,

$$c'_i(-) = \frac{\sum_{w \in E_i} \lambda_w v_w(-)}{\sum_{w \in E_i} \lambda_w}$$

Then

$$(c, c') = \sum_{w \in \mathcal{W}} \lambda_w (c, c')_w$$

After all, by (i), $c(-) = \sum_{w \in \mathcal{W}} \lambda_w v_w(-)$. And, by (ii), for all E_i in $\mathcal{E}$,

$$\left(\sum_{w \in E_i} \lambda_w \right) c'_i(-) = \sum_{w \in E_i} \lambda_w v_w(-)$$

so

$$c'_i(-) = \sum_{w \in E_i} \lambda_w v_w(-) + \sum_{w \notin E_i} \lambda_w c'_i(-)$$

[10] Recall: the convex hull of a set is the smallest convex set that includes that set.

as required.

Second, right-to-left. Suppose

$$(c, c') = \sum_{w \in \mathcal{W}} \lambda_w (c, c')_w$$

Then $c(-) = \sum_{w \in \mathcal{W}} \lambda_w c(-)$, which gives (i). And

$$c'_i(-) = \sum_{w \in E_i} \lambda_w v_w(-) + \sum_{w \notin E_i} \lambda_w c'_i(-)$$

which ensures that, if $\sum_{w \in E_i} \lambda_w > 0$, then

$$c'_i(-) = \frac{\sum_{w \in E_i} \lambda_w v_w(-)}{\sum_{w \in E_i} \lambda_w}$$

which gives (ii). That completes the proof. □

Theorem 5 *(c, c') is diachronically incoherent over $\mathcal{W}$ iff c' is not a superconditioning rule for c.*

Proof of Theorem 5. We will prove only the left-to-right direction; the other direction is straightforward. Suppose c' is not a superconditioning rule for c. Then (c, c') is not in the convex hull of $\{(c, c')_w : w \in \mathcal{W}\}$. So, by the Separating Hyperplane Theorem, there is a vector

$$S = \langle S_1, \ldots, S_n, S^1_1, \ldots, S^1_n, \ldots, S^k_1, \ldots, S^k_n \rangle$$

such that, for all w in $\mathcal{W}$,

$$S \cdot ((c, c')_w - (c, c')) < 0.$$

Now, if w is in E_i, then

$$\begin{aligned}
&(c, c')_w - (c, c') = \\
&\quad (v_w - c) \frown (c'_1 - c'_1) \frown \ldots \\
&\qquad \ldots \frown (c'_{i-1} - c'_{i-1}) \frown (w - c'_i) \frown (c'_{i+1} - c'_{i+1}) \frown \ldots \\
&\qquad\quad \ldots \frown (c'_n - c'_n) = \\
&\quad (v_w - c) \frown 0 \frown \ldots 0 \frown (v_w - c'_i) \frown 0 \frown \ldots \frown 0
\end{aligned}$$

So, for all w in E_i,

$$\begin{aligned}
& S \cdot ((c, c')_w - (c, c')) \\
= \; & (S_1, \ldots, S_n) \cdot (v_w - c) + (S^i_1, \ldots, S^i_n) \cdot (v_w - c'_i) \\
= \; & \sum_{j=1}^{n} S_j (v_w(X_j) - c(X_j)) + \sum_{j=1}^{n} S^i_j (v_w(X_j) - c(X_j)) \\
< \; & 0
\end{aligned}$$

Thus, at the earlier time, we offer each $\$S_j$-bet on X_j for $\$S_j c(X_j)$. And c demands you accept each at that price. And at the later time, if you learn E_i, we offer each $\$S^i_j$-bet on X_j for $\$S^i_j c'_i(X_j)$. And c'_i demands you accept each at that price. Now, the total payout of the bets at world w in E_i is:

$$\sum_{j=1}^{n} v_w(X_j)(S_j + S^i_j)$$

while the total price of the bets is

$$\sum_{j=1}^{n} S_j c(X_j) + \sum_{j=1}^{n} S^i_j c(X_j)$$

But, by the inequality above, the total price exceeds the total payout at every personally possible world. That completes the proof. □

Theorem 7 *Suppose $\mathfrak{I}$ is an additive strictly proper inaccuracy measure. Then (c, c') is accuracy dominated over $\mathcal{W}$ relative to $\mathfrak{I}$ iff c' is not a superconditioning rule for c.*

Proof of Theorem 7. Again, we prove only the left-to-right direction. Suppose $\mathfrak{I}$ is an additive, strictly proper inaccuracy measure. Then, by Proposition 2 in (Predd et al., 2009), there is a Bregman divergence $\mathfrak{D} : [0,1]^n \to [0,\infty]$ such that $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$. Then extend $\mathfrak{D}$ to be the Bregman divergence between $n(k+1)$-dimensional vectors that sums the Bregman divergences between the $k+1$ n-dimensional vectors. Thus,

$$\mathfrak{D}(b \frown b_1 \frown \ldots \frown b_k, c \frown c_1 \frown \ldots \frown c_k) = \mathfrak{D}(b, c) + \sum_{i=1}^{k} \mathfrak{D}(b_i, c_i)$$

Now suppose c' is not a superconditioning rule for c. Then (c, c') is not in the convex hull of $\{(c, c')_w : w \in \mathcal{W}\}$. So, by Proposition 3 of (Predd et al., 2009), there is $(c^\star, c^{\star\prime})$ such that, for all E_i in $\mathcal{E}$ and w in E_i,

$$\begin{aligned}&\mathfrak{D}(w \frown c'_1 \frown \ldots c'_{i-1} \frown w \frown c'_{i+1} \frown \ldots \frown c'_n, c^{\star\prime} \frown c^{\star\prime}_1 \frown \ldots \frown c^{\star\prime}_n)\\ &< \mathfrak{D}(w \frown c'_1 \frown \ldots c'_{i-1} \frown w \frown c'_{i+1} \frown \ldots \frown c'_n, c \frown c'_1 \frown \ldots \frown c'_n)\end{aligned}$$

But

$$\begin{aligned}&\mathfrak{D}(w \frown c'_1 \frown \ldots c'_{i-1} \frown w \frown c'_{i+1} \frown \ldots \frown c'_n, c \frown c'_1 \frown \ldots \frown c'_n) =\\ &\mathfrak{D}(w, c) + \mathfrak{D}(w, c'_i) = \mathfrak{I}((c, c'), w)\end{aligned}$$

And

$$\begin{aligned}&\mathfrak{D}(w \frown c'_1 \frown \ldots c'_{i-1} \frown w \frown c'_{i+1} \frown \ldots \frown c'_n, c^{\star\prime} \frown c^{\star\prime}_1 \frown \ldots \frown c^{\star\prime}_n) \geq\\ &\mathfrak{D}(w, c^\star) + \mathfrak{D}(w, c^{\star\prime}_i) = \mathfrak{I}((c^\star, c^{\star\prime}), w)\end{aligned}$$

So $\mathfrak{I}((c^\star, c^{\star\prime}), w) < \mathfrak{I}((c, c'), w)$ for all w in $\mathcal{W}$, as required. This completes the proof. □

References

Berto, F. (2014). On Conceiving the Inconsistent. *Proceedings of the Aristotelian Society*, *114*, 103–21.

Bjerring, J. C., & Skipper, M. (ta). Bayesianism for non-ideal agents. *Erkenntnis*.

Briggs, R. A., & Pettigrew, R. (2018). An accuracy-dominance argument for conditionalization. *Noûs*.

Brown, P. M. (1976). Conditionalization and expected utility. *Philosophy of Science*, *43*(3), 415–419.

Chalmers, D. J. (2011). Frege's Puzzle and the Objects of Credence. *Mind*, *120*(479), 587–635.

de Finetti, B. (1937 [1980]). Foresight: Its Logical Laws, Its Subjective Sources. In H. E. Kyburg, & H. E. K. Smokler (Eds.) *Studies in Subjective Probability*. Huntingdon, N. Y.: Robert E. Kreiger Publishing Co.

de Finetti, B. (1974). *Theory of Probability*, vol. I. New York: John Wiley & Sons.

Diaconis, P., & Zabell, S. L. (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, *77*(380), 822–830.

Dogramaci, S. (2018). Solving the Problem of Logical Omniscience. *Philosophical Issues (Supp. Noûs)*, *28*, 107–28.

Dorst, K. (2017). Lockeans Maximize Expected Accuracy. *Mind* 10.1093/mind/fzx028.

Earman, J. (1992). *Bayes Or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press.

Easwaran, K. (2016). Dr Truthlove, Or: How I Learned to Stop Worrying and Love Bayesian Probabilities. *Noûs*, *50*(4), 816–853.

Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese*, *140*, 97–119.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (Ed.) *Testing Scientific Theories (Midwest Studies in the Philosophy of Science, Vol. X)*, (pp. 99–131). Minneapolis: University of Minnesota Press.

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

Good, I. J. (1967). On the Principle of Total Evidence. *The British Journal for the Philosophy of Science*, *17*, 319–322.

Greaves, H., & Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, *115*(459), 607–632.

Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science*, *34*(4), 311–325.

Hájek, A. (2008). Dutch Book Arguments. In P. Anand, P. Pattanaik, & C. Puppe (Eds.) *The Oxford Handbook of Rational and Social Choice*, (pp. 173–195). Oxford: Oxford University Press.

Jago, M. (2012). Constructing Worlds. *Synthese*, *189*, 59–74.

Jago, M. (2014). *The Impossible: an Essay on Hyperintensionality*. Oxford, UK: Oxford University Press.

Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.

Jeffrey, R. (1992). *Probability and the Art of Judgment*. New York: Cambridge University Press.

Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, *65*(4), 575–603.

Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber, & C. Schmidt-Petri (Eds.) *Degrees of Belief*. Springer.

Konek, J. (2019). The Art of Learning. In T. S. Gendler, & J. Hawthorne (Eds.) *Oxford Studies in Epistemology*, vol. 7. Oxford University Press.

Lewis, D. (1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology*, (pp. 403–407). Cambridge, UK: Cambridge University Press.

Lewis, D. (2004). Letters to Priest and Beall. In B. Armour-Garb, J. C. Beall, & G. Priest (Eds.) *The Law of Non-Contradiction—New Philosophical Essays*, (pp. 176–7). Oxford, UK: Oxford University Press.

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

Pettigrew, R. (2020). *The Dutch Book Argument*. Elements in Decision Theory and Philosophy. Cambridge, UK: Cambridge University Press.

Predd, J., Seiringer, R., Lieb, E. H., Osherson, D., Poor, V., & Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions of Information Theory*, *55*(10), 4786–4792.

Priest, G. (1997). Sylvan's Box: A Short Story and Ten Morals. *Notre Dame Journal of Formal Logic*, *38*, 573–81.

Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford, UK: Oxford University Press.

Ramsey, F. P. (1926 [1931]). Truth and Probability. In R. B. Braithwaite (Ed.) *The Foundations of Mathematics and Other Logical Essays*, chap. VII, (pp. 156–198). London: Kegan, Paul, Trench, Trubner & Co.

Rothschild, D. (2019). Lockean Beliefs, Dutch Books, and Scoring Systems. *Review of Symbolic Logic*.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons.

Skyrms, B. (1987). Coherence. In N. Rescher (Ed.) *Scientific Inquiry in Philosophical Perspective*, (pp. 225–242). Pittsburgh: University of Pittsburgh Press.

Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese*, *192*(9), 2769–2793.

van Fraassen, B. (1980). Rational Belief and Probability Kinematics. *Philosophy of Science*, *47*, 165–187.

Vineberg, S. (2016). Dutch Book Arguments. In E. N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Weisberg, J. (2009). Commutativity or Holism? A Dilemma for Conditionalizers. *The British Journal for the Philosophy of Science*, *60*(4), 793–812.

Williams, J. R. G. (2018). Rational Illogicality. *Australasian Journal of Philosophy*, *96*(1), 127–141.

Wilson, J. (ms). Accuracy and Probability Kinematics. Unpublished manuscript.

Yagisawa, T. (1988). Beyond Possible Worlds. *Philosophical Studies*, *53*, 175–204.