Pettigrew, R., & Briggs, R. A. (2018). An Accuracy-Dominance Argument for Conditionalization. *Noûs*, *54*(1), 162-181. https://doi.org/10.1111/nous.12258

# An Accuracy-Dominance Argument for Conditionalization

January 23, 2018

**Abstract**

Epistemic decision theorists aim to justify Bayesian norms by arguing that these norms further the goal of epistemic accuracy — having beliefs that are as close as possible to the truth. The standard defense of Probabilism appeals to accuracy dominance: for every belief state that violates the probability calculus, there is some probabilistic belief state that is more accurate, come what may. The standard defense of Conditionalization, on the other hand, appeals to expected accuracy: before the evidence is in, one should expect to do better by conditionalizing than by following any other rule. We present a new argument for Conditionalization that appeals to accuracy-dominance, rather than expected accuracy. Our argument suggests that Conditionalization is a rule of coherence: failing to plan to conditionalize is not just a bad response to the evidence; it is also inconsistent.

## 1 Introduction

Epistemic decision theorists aim to justify Bayesian norms for credences by arguing that these norms further the goal of epistemic accuracy — having beliefs that are as close as possible to the truth. The standard defense of Probabilism, which says that your credences should obey the probability calculus, appeals to accuracy dominance: for every belief state that violates the norm, there is some belief state that satisfies it that is more accurate, come what may. The standard defense of Conditionalization, on the other hand, appeals to expected accuracy: before the evidence is in, one should expect to do better by conditionalizing than by following any other rule. We present a new argument for Conditionalization that appeals to accuracy-dominance, rather than expected accuracy. Our argument suggests that Conditionalization is a rule of coherence: failing to plan to conditionalize is not just a bad response to the evidence; it is also inconsistent.

We begin, in section 2, by describing the epistemic decision theory framework, and we present the standard accuracy-dominance justification for Probabilism. Then we present the standard expected-accuracy argument for Conditionalization in section 3 and describe some of its shortcomings in order to motivate our new accuracy-dominance argument, which we present in section 5. In between, in section 4, we compare the two standard pragmatic arguments for Conditionalization with the existing expected-accuracy argument and our new accuracy-dominance argument to show that they pair up naturally and to illuminate the structure of the accuracy-dominance argument. We conclude in section 6 and prove our main theorem in the Appendix.

## 2 The Accuracy-Dominance Argument for Probabilism

Epistemic decision theorists hold that partial belief aims at achieving *accuracy*, or closeness to the truth. In other words, belief aims at avoiding *inaccuracy*, or distance from the truth. Formally, we can characterise a measure of inaccuracy $\mathfrak{I}_s(c)$ as a function of two arguments:

**A credence function** $c$ whose inaccuracy is assessed. We follow the usual assumption that $c$ is defined over an algebra of propositions, or sets of worlds — we'll call this $\mathcal{A}$. And we'll assume that $\mathcal{A}$ is finite. We will not assume that $c$ is a probability function — only that its values lie between 0 and 1 inclusive.

**A state of the world** $s$ against which $c$'s inaccuracy is assessed. $c$'s inaccuracy depends partly on what $c$ says about the world, and partly on how the world turns out. A state of the world is a proposition which, for every proposition $A$ in $\mathcal{A}$, entails either $A$ or its negation. Thus, the states of the world form a partition on the set of worlds — in each world, exactly one state obtains. Indeed, they form the most fine-grained partition that $\mathcal{A}$ contains. We'll call the set of states $\mathcal{S}$. And sometimes we'll say that a proposition $A$ is true at a state $s$ if $s$ entails $A$, while $A$ is false at $s$ if $s$ does not entail $A$.

Among the many ways of measuring inaccuracy, one class of measures is often singled out for special interest. This is the class of *strictly proper inaccuracy measures*, which obey the following four constraints.

**Separability** There is a function $\mathfrak{i}$, which takes a state $s$ in $\mathcal{S}$, a credence function $c$ defined on $\mathcal{A}$, and a proposition $A$ in $\mathcal{A}$, such that

$$\mathfrak{I}_s(c) = \sum_{A \in \mathcal{A}} \mathfrak{i}_s(c, A).$$

We might think of $\mathfrak{i}_s(c, A)$ as giving the inaccuracy of the individual credence $c(A)$ in the state $s$.

**Continuity** $\mathfrak{i}_s(c, A)$ is a continuous function of $c(A)$.

**Extensionality** If $v_s(A) = v_{s'}(A')$ and $c(A) = c'(A')$, then $\mathfrak{i}_s(c, A) = \mathfrak{i}_{s'}(c', A')$ (where $v_s$ is the valuation function at the state of the world $s$ — that is, $v_s(A) = 1$, if $s$ entails $A$; and $v_s(A) = 0$, if $s$ entails $\neg A$).

**Strict Propriety** For every credence function $p$ defined on $\mathcal{A}$ that obeys the axioms of probability, and every credence function $q \neq p$ defined on $\mathcal{A}$,

$$\sum_{s \in \mathcal{S}} p(s)\mathfrak{I}_s(p) < \sum_{s \in \mathcal{S}} p(s)\mathfrak{I}_s(q)$$

Separability requires that the inaccuracy of a belief state be decomposable as a sum of the state's inaccuracies about the different propositions in the domain; this makes the inaccuracy measure additive. Continuity requires that the accuracy of an individual credence varies continuously with the credence. Extensionality requires that the inaccuracy at a state of the credence assigned to a proposition by a credence function is a function only of the truth value of that proposition at that state and the credence assigned to it. Strict Propriety requires that probability functions be 'smug' by assigning themselves higher expected accuracy than any other credence function.[1]

Here is the Brier score, the most famous of the proper scoring rules (but not the only one).

$$\mathfrak{I}_s(c) = \sum_{A \in \mathcal{A}} (v_s(A) - c(A))^2$$

[1]For arguments in favour of Strict Propriety, see [Gibbard, 2008, Joyce, 2009, Horowitz, 2014, Pettigrew, 2016, Konek, ms].

Using the assumption that the correct measure of inaccuracy is a proper scoring rule, epistemic decision theorists are able to defend many common norms. For instance, Predd et al. [2009], Joyce [2009], and Pettigrew [2016] argue for Probabilism, the view that credence functions should conform to the probability calculus, by appealing to *accuracy-dominance* considerations. The following properties play a central role in their argument.

**Strong Accuracy-Dominance** $c^*$ *strongly accuracy-dominates* $c$ iff

- for all states $s$,
  $\mathfrak{I}_s(c^*) < \mathfrak{I}_s(c)$

**Weak Accuracy-Dominance** $c^*$ *weakly accuracy-dominates* $c$ iff

- for all states $s$,
  $\mathfrak{I}_s(c^*) \leq \mathfrak{I}_s(c)$,
- and for some state $s$,
  $\mathfrak{I}_s(c^*) < \mathfrak{I}_s(c)$.

Someone who aims at accuracy should avoid credence functions that are even weakly accuracy-dominated. This gives us the following norm:

**Accuracy Dominance for Credence Functions** A believer is rationally required not to adopt a credence function $c$ if there is an alternative credence function $c^*$ such that (i) $c^*$ strongly accuracy-dominates $c$, and (ii) $c^*$ is not itself even weakly accuracy-dominated by any other credence function.

Predd et al. [2009] show that Probabilism is necessary and sufficient for avoiding accuracy-dominance: every non-probability function is strongly accuracy-dominated by a probability function, while no probability function is even weakly accuracy-dominated by any other credence function. Thus, if we measure inaccuracy using a strictly proper inaccuracy measure, then Accuracy Dominance for Credence Functions entails Probabilism.

Arguments for norms other than Probabilism typically do not appeal to dominance reasoning. The epistemic utility argument for the Principle of Indifference appeals to minimax reasoning [Pettigrew, 2014]; while arguments for other norms — including Conditionalization [Greaves and Wallace, 2006, Easwaran, 2013, Leitgeb and Pettigrew, 2010], reflection [Easwaran, 2013], conglomerability [Easwaran, 2013], and the Principal Principle [Pettigrew, 2013] — appeal to considerations of *expected* accuracy. In expected accuracy arguments, the idea is that, from the standpoint of some particular probability function, obeying the norm is a better epistemic bet than violating it. Unlike accuracy-dominance, expected accuracy is defined only relative to a probability function.

It is often claimed that failure to conditionalize is a form of logical inconsistency over time [Armendt, 1992] [Christensen, 1991, 1996], [Lewis, 2010]. (Much of the subsequent debate centers on whether there is anything wrong with logical inconsistency over time.) In light of expected accuracy arguments, this claim about logical consistency is puzzling. It may be unreasonable to take a lousy bet, but there is nothing logically inconsistent about doing so — for all anyone knows, you might win a fortune at the casino, and your scratch lotto ticket might pay off.

An accuracy-dominance argument for Conditionalization, on the other hand, would suggest that failure to conditionalize is a form of logical inconsistency.[2] Just as there is something inconsistent about preferences that leave a person vulnerable to a sure monetary loss (independently of how contingent events turn out), there is something inconsistent about a belief-like state that leaves a person vulnerable to a sure loss of accuracy (independent of how contingent events turn out). To be inconsistent is to set oneself for epistemic failure, come what may.

A way of understanding the difference between inconsistency and mere unreasonableness is that norms of consistency take wide scope, while requirements of reasonableness take narrow scope. If Conditionalization is a requirement of consistency, then what you ought to do is: adopt a prior credence function and a plan for updating on your evidence, such that your planned later credence is guaranteed to be equal to your earlier credence conditional on subsequent evidence. If Conditionalization is a requirement of reasonableness, on the other hand, what you ought to do *when you have a particular prior credence function* is: plan to conditionalize on that prior credence function.

A few caveats are in order. First, Conditionalization may be both a requirement of consistency (when read in wide-scope form) and a requirement of reasonableness (when read narrow-scope form). So while our conclusion can't be established by existing arguments for Conditionalization, nothing in our argument shows that the existing arguments are unsound. Second, even if Conditionalization is a requirement of consistency, it may sometimes be rational to be inconsistent and violate Conditionalization. As Vineberg [1997] points out, a flawed agent who can't help holding an unreasonable attitude may have to choose between making her attitudes consistent with the unreason-

[2] Vineberg [2001] makes an exactly analogous point about accuracy-dominance arguments for Probabilism.

able attitude, and thereby more unreasonable, or making her other attitudes reasonable, but thereby more inconsistent because of the unreasonable attitude's fixity. Since consistency may not always trump reasonableness, consistency may not always be rationally required.

Still, there is value in establishing that Conditionalization is (among other things) a requirement of consistency. Thinking of Conditionalization this way can provide insight into what the norm requires of us (something wide-scope), why we should plan to conditionalize (for many of the same reasons that we should conform to the probability axioms), and when the norm of Conditionalization is trumped by other requirements (in cases where reasonableness is more important than consistency, and we can't have both).

In the next section, we will consider a representative expected-accuracy argument for Conditionalization, by Greaves and Wallace [2006]. We will then adapt key elements of the authors' framework to create a new argument for Conditionalization — this time, an accuracy-dominance argument.

## 3 Greaves and Wallace's Expected-Accuracy Argument for Conditionalization

Greaves and Wallace [2006] begin by assuming that every agent has

**a prior credence function** $c$**,** held at time $t_0$, and assumed to be a probability function defined on $\mathcal{A}$, and

**an evidence partition** $\mathcal{E} = \{E_1, \ldots, E_n\}$ where each member of $\mathcal{E}$ is a proposition in $\mathcal{A}$ that the believer might learn between $t_0$ and a later time $t_1$.

Agents are then meant to choose among the range of

**credal acts,** or functions $a$ that map each member of $E$ of $\mathcal{E}$ to a probability function $a_E$ defined on $\mathcal{A}$.[3]

On the intended interpretation, each credal act is a plan about which credence function to adopt at $t_1$ in light of evidence received between $t_0$ and $t_1$. The plan may call for different credences to be adopted depending on which evidence is received. The plan represented by the function $a$ tells the believer to adopt credence function $a_E$ if she learns that $E$.

[3]We use "credal acts" to refer to what Greaves and Wallace call "available credal acts". They also consider a more fine-grained model in which credal acts are functions from *states* to probability functions. The objects that

Within this framework, Greaves and Wallace give the following definition:

**A conditionalization plan for credence function** $c$ is an act that maps each evidence proposition $E$ in $\mathcal{E}$ to $c(\cdot|E)$, whenever $c(E) > 0$. That is, $a$ is a conditionalization plan for $c$ if $c(A \cap E) = c(E)a_E(A)$ for all propositions $A$ on which $c$ is defined.

Next, they extend the definition of inaccuracy. Not only can we measure the inaccuracy of a credence function at a state of the world; we can also measure the inaccuracy of a credal act at a state of the world. Remember: $\mathcal{E}$ is a partition. So, every state of the world $s$ entails exactly one evidence proposition in $\mathcal{E}$, which we might write as $E_s$. Therefore, a believer who performs a credal act $a$ at state $s$ will end up adopting $a_{E_s}$, the credence function that $a$ assigns to the evidence proposition $E_s$ that is true at $s$. So the accuracy of act $a$ in state $s$ is just the accuracy of $a_{E_s}$ — that is, $\mathfrak{I}_s(a) := \mathfrak{I}_s(a_{E_s})$. The measure of inaccuracy for credence functions thus uniquely determines the measure of accuracy for credal acts.

Using two additional assumptions, Greaves and Wallace argue that believers are rationally required to conditionalize. The first of these assumptions is Strict Propriety. The second assumption is

**Minimize Expected Inaccuracy** A believer with credence function $c$ is rationally required to choose a credal act that minimizes the *expected* degree of inaccuracy from the vantage point of $c$. Given a credal act $a$, its expected inaccuracy from the vantage point of $c$ is defined as follows:

$$\exp_c(\mathfrak{I}(a)) = \sum_{s \in \mathcal{S}} c(s)\mathfrak{I}_s(a)$$

Thus, a believer with credence function $c$ is rationally required to choose a credal act $a^*$ such that, for all credal acts $a$,

$$\exp_c(\mathfrak{I}(a^*)) \leq \exp_c(\mathfrak{I}(a))$$

From Strict Propriety, Greaves and Wallace derive two consequences: first, every conditionalization plan on $c$ has the same expected inaccuracy from the vantage point of $c$;

we are calling "credal acts" can be embedded in the fine-grained model, but the fine-grained model contains additional *unavailable* credal acts. Since the unavailable credal acts do not correspond to viable epistemic plans, we will henceforth ignore them.

second, from the vantage point of $c$, every conditionalization plan on $c$ has strictly lower expected inaccuracy than every credal act that isn't a conditionalization plan on $c$. Thus, given the norm Minimize Expected Inaccuracy, believers with probabilistic credences who are choosing between credal acts are rationally required to pick a conditionalization plan on their credence function. That is, together with Strict Propriety, Minimize Expected Inaccuracy establishes a narrow scope version of the synchronic norm that Kenny Easwaran [2013] calls Plan Conditionalization:

**Plan Conditionalization (narrow scope)** If you have a credence function $c$ at $t_0$, and you know that between $t_0$ and $t_1$ you will learn which member of the partition $\mathcal{E}$ is true (and nothing more), you should, at $t_0$, plan to update by conditionalization when the evidence comes in at $t_1$.

But there is another narrow scope version of Conditionalization that is perhaps more often stated. This version is genuinely diachronic.

**Diachronic Conditionalization (narrow scope)** If you have a credence function $c$ at $t_0$, and between $t_0$ and $t_1$ you learn $E$ and nothing more, then at $t_1$ you should adopt $c(-|E)$ as your credence function.

Does the Greaves and Wallace argument also establish this? There are two ways in which it might. First: you might think that it is irrational to make a plan at an earlier time to do something at a later time and then at that later time not to follow through on that plan, at least when nothing occurs in the interim that you did not account for as a possibility when you were making the original plan. This is the principle that Sarah Paul [2014] calls *Diachronic Continence*. Together, Plan Conditionalization (narrow scope) and Diachronic Continence entail Diachronic Conditionalization (narrow scope).

Second: Minimize Expected Inaccuracy says that an agent should pick a credal act that has minimal expected inaccuracy from the vantage point of their current credence function. Now, suppose that you have credence function $c$ at $t_0$ and you learn $E$ and nothing more between $t_0$ and $t_1$. You are now at $t_1$, and you are deciding how to respond to learning $E$. From the vantage point of which credence function should you make this decision? If it is still $c$, then the Greaves and Wallace argument says that you should pick a conditionalizing plan for $c$ at $t_1$, and given that you also know $E$, you should then adopt its recommendation and adopt $c(-|E)$. Thus, if we are obliged to take your credence function $c$ at $t_0$ also to be the credence function from whose vantage point you must choose between credal acts at $t_1$, then we can also establish Diachronic Conditionalization (narrow scope). But it isn't clear that we are so obliged. After all, at $t_1$, we have evidence $E$ and yet typically our credence function $c$ from $t_0$ will not encode that evidence. Thus, $c$ is flawed in a certain way. And, given this flaw, it isn't obvious that we must still turn to $c$ for recommendations at $t_1$.

In sum: if we grant Minimize Expected Inaccuracy, Greaves and Wallace's argument certainly establishes Plan Conditionalization (narrow scope); and, if supplemented with further, controversial assumptions, it establishes Diachronic Conditionalization (narrow scope). But this raises the question: should we grant Minimize Expected Inaccuracy? We might wonder: what's so special about $c$ that it gets to dictate the credal act that we should adopt even at $t_0$?

Perhaps $c$ is special because it is the credence function uniquely supported by the evidence. But in that case, Greaves and Wallace have not shown that Conditionalization is a norm governing all partial believers. They have shown that partial believers whose credence functions are uniquely best supported by the evidence ought to conditionalize.

Or perhaps $c$ is special because of the relation that believers bear to their own credence functions. If you choose an act with lower expected value over an act with higher expected value, you are being instrumentally irrational. And if the value in question is epistemic, then you are being epistemically irrational. This version of the argument shows that there is something epistemically foolhardy about failing to conditionalize, but it doesn't establish that Conditionalization is a norm of logical consistency. It's perfectly consistent to perform an act with low expected utility, like playing the lottery or pursuing the career in arts; for many people, it even ends well.[4] We propose that we can do better: we can establish that violating Conditionalization is not just a bad idea, but inconsistent. We can do this using an accuracy-dominance argument, which does not rely on assuming anything about a particular initial credence function $c$.

Another concern about Minimize Expected Inaccuracy is that, even as a principle of reasonableness, let alone as a norm of consistency, it is controversial. While all decision theorists accept Dominance — at least in cases in which the acts to be evaluated are inde-

[4][Redacted for review] suggests that we might supplement the second version of the expected accuracy dominance argument with the observation that, no matter what beliefs you adopt consistent with the probability calculus, planning to do anything other than conditionalize is foolhardy by your own lights. Perhaps this argument does establish Conditionalization as a norm of consistency. Our accuracy-dominance argument operates by an interestingly different mechanism.

pendent of the states at which they are evaluated — many decision theorists reject principles like Minimize Expected Inaccuracy. These are the non-expected utility theorists, and they include those who would replace expected utility maximization as a norm by something like prospect theory [Kahneman and Tversky, 1979] or rank-dependent utility theory [Quiggin, 1982] or risk-weighted expected utility theory [Buchak, 2013] in which Minimize Expected Inaccuracy is replaced by a different norm. They are motivated by cases like the Allais paradox, in which people report preferences that cannot be rationalized by expected utility theory. Such preferences involve a level of risk aversion that cannot be accommodated in expected utility theory by tweaking the utility function. But if Minimize Expected Inaccuracy is controversial and Dominance is not, then surely an argument for Conditionalization that assumes only the latter should be preferred over one that assumes the former.

## 4 Pragmatic and epistemic arguments for Conditionalization

Before we present our accuracy-dominance argument for Conditionalization, it will be helpful to explain how the accuracy arguments for this norm — Greaves and Wallace's expected-accuracy argument and our new accuracy-dominance argument — relate to the pragmatic arguments in its favour. There are two pragmatic arguments for Conditionalization, and they correspond naturally to the two accuracy arguments just mentioned.

The first, which is less well known, is an expected-utility argument due to Peter M. Brown [1976]. Like Greaves and Wallace, Brown assumes that we have a prior probabilistic credence function $c$ on $\mathcal{A}$, and an evidence partition $\mathcal{E}$. Also like Greaves and Wallace, he wishes to assess each candidate credal act defined on $\mathcal{E}$ from the vantage point of $c$; and he wishes to do so using expected value. However, where Greaves and Wallace take the value of a credal act to be its epistemic value and, in particular, its accuracy, Brown takes its value to be its pragmatic value and, in particular, its utility. Now, what is the pragmatic value, or utility, of a credal act? Brown assumes that there is a range of options between which you must choose. Each of these options is assigned a utility at each state of the world. And thus, for any probabilistic credence function, there is a subset of this range of options that contains the ones that maximise expected utility from the vantage point of that credence function. Let's assume, furthermore, that our agent has a tie-breaker function. So, if there is more than one option that maximises expected utility, the agent uses her tie-breaker function to select exactly one of them as her choice. This is the option that the agent would pick were she to have that credence function and were she to be faced with that range of options. Brown then takes the pragmatic value of a credal act at a state to be the utility of the option that the agent would pick were she to receive as evidence whichever evidence proposition is true at that state. Thus, the pragmatic value of a credal act is the value of the option it would lead you to choose if you were to make your choice once the evidence is in and you have updated in line with that credal act. Then Brown shows that, for any range of options, planning to conditionalize on your current credence function when you obtain new evidence maximizes your expected utility. What's more, for any credence function $c$, any evidence partition $\mathcal{E}$, and any credal act other than conditionalization, there is a range of options such that, if you must choose between those options after your evidence comes in from that partition, then you expect conditionalizing to do strictly better than the alternative. Therefore, you should plan to conditionalize.

Clearly, this argument is closely related to the Greaves and Wallace argument. It uses the same decision-theoretic norm — namely, Maximize Expected Utility — to choose between credal acts on the basis of a prior credence function. It differs from the Greaves and Wallace argument in the sort of value to which it appeals — Greaves and Wallace appeal to epistemic value, or accuracy; Brown appeals to pragmatic value, or utility. For Greaves and Wallace, the value of a credal act at a state is the accuracy of the credence function it recommends at that state; for Brown, the value of a credal act at a state is the utility of the option that you would choose were you to adopt the credence function it recommends at that state.

The other, much better known pragmatic argument for Conditionalization is the so-called *diachronic Dutch book argument* [Lewis, 2010]. This runs as follows:

First, some terminology: A *credal strategy* is a pair consisting of a credence function $c$ defined on $\mathcal{A}$ and a credal act $a$ defined on an evidence partition $\mathcal{E}$, where each $a_E$ is defined on $\mathcal{A}$. On our intended interpretation, a credal strategy is a two-stage plan that specifies which credence function the believer will adopt at $t_0$, before the evidence is in, and which credence function she will adopt at $t_1$, after she has learned which $E$ in $\mathcal{E}$ is true. We say that $\langle c, a \rangle$ is a *probabilistic strategy* if $c$ is a probability function, while $\langle c, a \rangle$ is a *conditionalizing strategy* if $a$ is a conditionalizing plan for $c$.

Now, as with all Dutch book arguments, this one makes an assumption about which bets you will accept, given your credences: if you have credence $p$ in proposition $A$, then you will pay £$Sp$ for a bet that returns £$S$ if $A$ is true and £0 if $A$ is false. Granted this, we can show that, if your credal strategy is $\langle c, a\rangle$ and $\langle c, a\rangle$ is not a probabilistic and conditionalizing strategy, then there are books of bets $B$ and $B_E$ for each $E$ in $\mathcal{E}$ such that (i) you will accept each of the bets in $B$ now; (ii) you will accept each of the bets in $B_E$ if you learn $E$ and update your credence function to $a_E$ in accordance with $a$; and (iii) whichever $E$ you learn, it is guaranteed that $B$ and $B_E$ will together lose you money. Thus, your credal strategy leads you to choose a strongly dominated sequence of options — if you had chosen instead to refuse all bets in $B$ and $B_E$, you would have been guaranteed to do better. And it seems natural to say that, just as choosing a dominated option is irrational, so it is irrational to adopt a credal strategy that sanctions choosing such an option.

This is the pragmatic argument for Conditionalization that is most closely analogous to the accuracy-dominance argument that we wish to present. It uses the same decision-theoretic norm, namely, Dominance. It differs from our argument because it appeals to pragmatic value rather than epistemic value. It shows that credal strategies that are not probabilistic and conditionalizing sanction certain betting behaviour that, from the point of view of pragmatic value, or utility, is dominated by refraining from betting. Our argument, in contrast, shows that credal strategies that are not probabilistic and conditionalizing are, from the point of view of epistemic value, or accuracy, dominated by alternative strategies.

## 5 An Accuracy-Dominance Argument for Conditionalization

Like Greaves and Wallace, we will assume that each agent is endowed with a set $\mathcal{A}$ of propositions to which she assigns credences, a corresponding set $\mathcal{S}$ of epistemically possible states of the world, and an evidence partition $\mathcal{E} \subseteq \mathcal{A}$. However, we will not assume that each agent is endowed with a fixed initial credence function, or that she chooses among epistemic acts. Instead, we assume that she chooses among the set of possible credal strategies.

The key conceptual move is to switch from measuring the inaccuracy of credal acts, at states, to measuring the inaccuracy of credal strategies, also at states. This is a larger conceptual step than the switch from measuring the inaccuracy of credence functions to measuring the inaccuracy of credal acts. Given a way of measuring inaccuracy for credence functions, we could pin down a unique right way of measuring inaccuracy for credal acts. But to pin down a unique right way of measuring inaccuracy for credal strategies, we need one more assumption:

**Temporal Separability** $\mathfrak{I}_s(\langle c, a \rangle) = \mathfrak{I}_s(c) + \mathfrak{I}_s(a) \quad (= \mathfrak{I}_s(c) + \mathfrak{I}_s(a_{E_s}))$

Temporal Separability generalizes Separability; it says that the inaccuracy of a strategy, i.e., a pair of credal acts, is the sum of the inaccuracies of the two acts in the strategy.[5] Once we assume Temporal Separability, any way of measuring inaccuracy for individual credence functions uniquely determines the inaccuracy of any pair of credence functions at a state, and therefore uniquely determines the inaccuracy of any credal strategy at any state.

Given the assumptions of Separability, Continuity, Extensionality, Strict Propriety, and Temporal Separability, we can prove the following:

**Theorem 1** Let $\mathfrak{I}$ be a measure of inaccuracy satisfying Separability, Continuity, Extensionality, Strict Propriety, and Temporal Separability. Then

(I) For each credal strategy that *is not* probabilistic, *there is* an alternative credal strategy that is probabilistic and conditionalizing that *weakly* accuracy-dominates it.

(II) For each credal strategy that *is not* conditionalizing, there is an alternative credal strategy that is probabilistic and conditionalizing that *strongly* accuracy-dominates it.

(III) For each credal strategy that *is* probabilistic and conditionalizing, *there is no* alternative credal strategy whatsoever that even *weakly* accuracy-dominates it.

[5] In fact, our argument would go through if we took the inaccuracy of a credal strategy to be the *weighted* sum of the inaccuracy of its components: that is,

**Temporal Separability**$^{-}$ For $\alpha, \beta > 0$, $\mathfrak{I}_s(\langle c, a \rangle) = \alpha\mathfrak{I}_s(c) + \beta\mathfrak{I}_s(a)$

Doing this would allow us to accommodate some degree of temporal discounting for epistemic choices, for instance — if I care less about my future accuracy than my current accuracy, then I pick $\beta < \alpha$.

Now, suppose we adapt Accuracy Dominance for Credence Functions so that it applies instead to credal strategies:

**Accuracy Dominance for Credal Strategies** A believer is rationally required not to adopt a credal strategy $\langle c, a\rangle$ if there is an alternative credal strategy $\langle c^*, a^*\rangle$ such that (i) $\langle c^*, a^*\rangle$ strongly accuracy-dominates $\langle c, a\rangle$, and (ii) $\langle c^*, a^*\rangle$ is not itself even weakly accuracy-dominated by any other credal strategy.

Then Theorem 1 shows that, if we measure inaccuracy using strictly proper inaccuracy measures, then Accuracy Dominance for Credal Strategies entails at least the wide scope version of Plan Conditionalization:

**Plan Conditionalization (wide scope)** If you know that between $t_0$ and $t_1$ you will learn which member of the partition $\mathcal{E}$ is true (and nothing more), you should, at $t_0$, adopt a credence function $c$ and a credal act $a$ such that $a$ is a conditionalizing plan for $c$.

As with Greaves and Wallace's expected-accuracy argument, we might ask whether our accuracy-dominance argument establishes Plan Conditionalization (wide scope) only or also Diachronic Conditionalization (wide scope) as well, or even the more familiar Diachronic Conditionalization (narrow scope).

**Diachronic Conditionalization (wide scope)** if between $t_0$ and $t_1$ you learn $E$ and nothing more, you should have a credence function $c_0$ at $t_0$ and a credence function $c_1$ at $t_1$ such that $c_1 = c_0(-|E)$.

At this point, skeptics about Conditionalization may look askance. Should credal strategies be subject to rational evaluation? Why should rationality require anyone to be consistent *over time*? Someone who was consistent yesterday, and is consistent today, has done enough to satisfy the demands of rationality.

But notice: the argument doesn't just assume that believers should be consistent over time; it provides a reason for consistency over time. At any given time, it is better to have accurate beliefs than to have inaccurate ones. Furthermore, it is better for a person to be more accurate over the course of a lifetime (in total or on average), than it is for that person to be less accurate. The accuracy-dominance argument for Conditionalization justifies the wide scope norm of diachronic consistency by appealing to the value of diachronic accuracy.

Of course, skeptics could still dig their heels in, and insist that what matters is accuracy at a time — that accuracy across time is irrelevant. Or they could argue that, while being accuracy-dominated across time is unfortunate, it is a misfortune that no one can be blamed for — since the problem is not the believer's credence function at any one time, but rather, a global property of her behavior across time. We are not sure how to adjudicate this debate with the skeptic. The accuracy-dominance argument is an improvement on arguments that rely on the claim that rational beings are subject to diachronic norms of consistency, but it is not yet a proof of Conditionalization based on self-evident premises. It shows, at least, where the controversy should lie.

Can we infer from Diachronic Conditionalization (wide scope) to Diachronic Conditionalization (narrow scope), the more familiar version of the Conditionalization norm? We think not, and for the same reason that you can't infer from a wide scope diachronic norm in the practical case to the corresponding narrow scope diachronic norm. Suppose that, on Saturday morning, I buy tickets to go to the cinema on Saturday evening. Then, on Saturday afternoon, I think better of it, and I sell my tickets to a friend at slightly less than I paid for them. Then there is a alternative strategy that dominates my actual strategy. If I were to refrain from buying the tickets on the Saturday morning, and then to refrain from selling them in the afternoon, I would end up with more money overall than I end up with given my actual pair of actions, which guarantees me a loss. But that doesn't mean that, if I buy the tickets on Saturday morning, it is irrational for me to sell them at a reduced price on Saturday afternoon. I bought the tickets because that was the best choice by my lights in the morning; and I sold them again in the afternoon because that was the best choice by my lights in the afternoon [Moss, 2015]. Between the morning and the afternoon, I had changed my mind. And there is nothing irrational about changing your mind. The same holds in the epistemic case. Thus, our argument certainly establishes Plan Conditionalization (wide scope). And we submit that it also establishes Diachronic Conditionalization (wide scope). But it does not establish Diachronic Conditionalization (narrow scope).

## 6 Conclusion

We have provided a new and illuminating argument that Conditionalization governs the correct response to new evidence. While Conditionalization itself is widely accepted, the reasons for conditionalizing are still subject to dispute. It is more than merely a pragmatic norm — the reasons for conditionalizing go beyond the fact that it is profitable. But philosophers disagree about the deeper reasons to conditionalize. Our argument provides an answer: failing to plan to conditionalize involves an inconsistency between a believer's credences, on the one hand, and her dispositions to change her mind in response to new evidence, on the other. Any strategy that conflicts with Conditionalization is self-defeating: it results in achieving lower accuracy than an available alternative, come what may.

The key move in the argument was to generalize the concept of accuracy twice over. First, we took advantage of Greaves and Wallace's idea of measuring the accuracy of credal acts, in addition to credence functions. Next, we generalized the concept of accuracy to apply to strategies, or sequences of credal acts, in addition to individual acts. Together, these steps allowed us to prove an accuracy-dominance theorem for Conditionalization, using the assumption that we measure inaccuracy using a strictly proper inaccuracy measure.

Armed with these new concepts, it is possible to show not just that Conditionalization is a norm of rationality, but also something about why. Conditionalization is not just a good idea — it's a law of coherent updating.

# References


Brad Armendt. Dutch strategies for diachronic rules: When believers see the sure loss coming. *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association*, Volume One: Contributed Papers:217–229, 1992.

A Banerjee, X Guo, and Hui Wang. On the optimality of conditional expectation as a Bregman predictor. *Information Theory, IEEE Transactions on*, 51(7):2664–2669, July 2005.

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

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

David Christensen. Clever Bookies and Coherent Beliefs. *The Philosophical Review*, 100 (2):229–247, April 1991.

David Christensen. Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers. *The Journal of Philosophy*, 93(9):450, September 1996.

Bruno de Finetti. *Theory of Probability*, volume 1. Wiley, New York, 1974.

Kenny Easwaran. Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science*, 80(1):119–142, January 2013.

Allan Gibbard. Rational Credence and the Value of Truth. In T. Gendler and J. Hawthorne, editors, *Oxford Studies in Epistemology*, volume 2. Oxford University Press, 2008.

Hilary Greaves and David Wallace. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632, July 2006.

Sophie Horowitz. Immoderately rational. *Philosophical Studies*, 167:41–56, 2014.

James M Joyce. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In *Degrees of Belief*, pages 263–297. Springer Netherlands, Dordrecht, January 2009.

Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263, 1979.

Jason Konek. A Veritist's Guide to Strict Propriety. Unpublished manuscript, ms.

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

David Lewis. Why Conditionalize? In Antony Eagle, editor, *Philosophy of Probability*, pages 403–407. Routledge, 2010.

Sarah Moss. Credal Dilemmas. *Noûs*, 49(4):665–683, 2015.

Sarah K. Paul. Diachronic Incontinence is a Problem in Moral Philosophy. *Inquiry: An Interdisciplinary Journal of Philosophy*, 57(3):337–355, 2014.

Richard Pettigrew. A New Epistemic Utility Argument for the Principal Principle. *Episteme*, 10(1):19–35, March 2013.

Richard Pettigrew. Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*, 90(1), 2014.

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

J B Predd, R Seiringer, and E H Lieb. Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(4):4786–4792, 2009.

John Quiggin. A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization*, 3:323–43, 1982.

Susan Vineberg. Dutch Books, Dutch Strategies and What They Show about Rationality. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 86 (2):185–201, May 1997.

Susan Vineberg. The notion of consistency for partial belief. *Philosophical Studies*, 2001.

## 7 Appendix: Proof of Theorem 1

Recall our

**Theorem 1** Let $\mathfrak{I}$ be a measure of inaccuracy satisfying Separability, Continuity, Extensionality, Strict Propriety, and Temporal Separability. Then

(I) For each credal strategy that *is not* probabilistic, *there is* an alternative credal strategy that is probabilistic and conditionalizing that *weakly* accuracy-dominates it.

(II) For each credal strategy that *is not* conditionalizing, there is an alternative credal strategy that is probabilistic and conditionalizing that *strongly* accuracy-dominates it.

(III) For each credal strategy that *is* probabilistic and conditionalizing, *there is no* alternative credal strategy whatsoever that even *weakly* accuracy-dominates it.

Our argument adapts and generalizes the proof in [de Finetti, 1974].

### 7.1 Proof of Theorem 1(I)

Fix the algebra $\mathcal{A} = \{X_1, \ldots, X_m\}$ on which our agent's credence functions are defined; and fix the partition $\mathcal{E} = \{E_1, \ldots, E_n\}$ on which her credal acts are defined. Thus, suppose $c$ is defined on $\mathcal{A}$ and $a$ is defined on $\mathcal{E}$. If $\langle c, a\rangle$ is non-probabilistic, then we can find a probabilistic strategy that weakly accuracy-dominates it. By Predd et. al.'s result, for each credence function amongst $c$, $a_{E_1}, \ldots, a_{E_n}$ that is not a probability function, there is an alternative credence function that strongly accuracy-dominates it. If we replace each such non-probabilistic credence function with one that strongly accuracy-dominates it, then the resulting strategy weakly accuracy-dominates $\langle c, a\rangle$, since it does better in all states in which some $E_i$ holds for which we replaced $a_{E_i}$; and in all states, if we replaced $c$. QED.

### 7.2 Proof of Theorem 1(II)

We begin by proving something that appears weaker than (II):

(II′) For each credal strategy that *is not* conditionalizing, there is an alternative credal strategy that *strongly* accuracy-dominates it.

Then, later, we will show that at least one of the weakly (or strongly) accuracy-dominating strategies must be probabilistic and conditionalizing.

We proceed in three steps.

1. Reformat strategies and states.
2. Find a particular accuracy-dominating strategy $\langle c^*, a^*\rangle$ (where $c^*$ is also defined on $\mathcal{A}$ and $a^*$ is also defined on $\mathcal{E}$).
3. Show that $\langle c^*, a^*\rangle$ accuracy-dominates $\langle c, a\rangle$.

**Reformatting Strategies** The first step is to rewrite each strategy and each state as a single vector with the same number of places. Suppose $\langle c, a\rangle$ is a credal strategy where the domain of $c$ and each $a_{E_i}$ is $\mathcal{A} = \{X_1, \ldots X_m\}$. Then we can create a vector $\vec{c}$ by concatenating $c$ with each of the $a_{E_i}$s, like so.

$$\vec{c} = c \frown a_{E_1} \frown \ldots \frown a_{E_n}$$

In other words,

$$\vec{c} = \langle \underbrace{c(X_1), \ldots c(X_m)}_{c},\ \underbrace{a_{E_1}(X_1), \ldots, a_{E_1}(X_m)}_{a_{E_1}},\ \ldots,\ \underbrace{a_{E_n}(X_1), \ldots, a_{E_n}(X_m)}_{a_{E_n}} \rangle$$

**Reformatting States** Suppose $\langle c, a \rangle$ is a strategy and $s$ is a state that entails $E_j$. Let $v_s$ be the valuation function for the state $s$: that is, $v_s(X_i) = 1$ if $s$ entails $X_i$; and $v_s(X_i) = 0$ if $s$ entails $\overline{X_i}$. Then we can define a vector $\vec{c}_s$ by taking the $\vec{c}$ generated by $\langle c, a \rangle$, and replacing both $c(X_i)$ and $c_{E_j}(X_i)$ with $v_s(X_i)$, for each $X_i$ in $\mathcal{A}$. This gives us:

$$\vec{c}_s = v_s \frown c_{E_1} \frown \ldots c_{E_{j-1}} \frown v_s \frown c_{E_{j+1}} \ldots \frown c_{E_n}$$

In other words,

$$\begin{aligned} \vec{c}_s = \langle & \underbrace{v_s(X_0), \ldots v_s(X_m)}_{v_s}, \underbrace{c_{E_1}(X_0), \ldots, c_{E_1}(X_m)}_{c_{E_1}}, \ldots, \\ & \underbrace{c_{E_{j-1}}(X_0), \ldots, c_{E_{j-1}}(X_m)}_{c_{E_{j-1}}}, \underbrace{v_s(X_0), \ldots, v_s(X_m)}_{v_s}, \\ & \underbrace{c_{E_{j+1}}(X_0), \ldots, c_{E_{j+1}}(X_m)}_{E_{j+1}}, \ldots, \underbrace{c_{E_n}(X_0), \ldots, c_{E_n}(X_m)}_{c_{E_n}} \rangle \end{aligned}$$

**Characterizing Conditionalization** A crucial component of Predd, et al.'s proof that all and only the probabilistic credence functions are non-dominated is a geometric characterization of those credence functions. They show that $c$ is a probability function iff $c$ is in the convex hull of the $v_s$s. That is, the set of probability function is precisely the convex hull of $\{v_s : s \in \mathcal{S}\}$. By reformatting the strategies and states as we have just down, we can give a similarly useful alternative characterization of conditionalizing strategies.

**Lemma 3** Given that $\vec{c} = \langle c, a \rangle$ is a probabilistic strategy, $\vec{c}$ is a conditionalizing strategy iff $\vec{c}$ is in the convex hull of $\{\vec{c}_s : s \in \mathcal{S}\}$.

Proof of Lemma 3:

**(Left-to-right)** To show the left-to-right direction, we suppose that $\vec{c} = \langle c, a \rangle$ is a conditionalizing strategy, and show that $\vec{c}$ is in the convex hull of $\{\vec{c}_s : s \in \mathcal{S}\}$. That is, we show that there are non-negative real numbers $\lambda_s$ (for $s \in \mathcal{S}$) such that $\vec{c} = \sum_s \lambda_s \vec{c}_s$.

1. Let $\lambda_s = c(s)$

2. $c = \sum_s \lambda_s v_s$

   (By 1, and the assumption that $c$ is a probability function)

3. Since $\langle c, a \rangle$ is a conditionalizing strategy, for each $E_i \in \mathcal{E}$ and $X \in \mathcal{A}$,

$$c(E_i)a_{E_i}(X) = c(X \cap E_i)$$

   Thus:

$$\begin{aligned} a_{E_i}(X) &= a_{E_i}(X) - c(E_i)a_{E_i}(X) + c(X \cap E_i) \\ &= (1 - c(E_i))a_{E_i}(X) + c(X \cap E_i) \\ &= \sum_{s \not\subseteq E_i} c(s)a_{E_i}(X) + \sum_{s \subseteq E_i} c(s)v_s(X) \\ &= \sum_{s \not\subseteq E_i} \lambda_s a_{E_i}(X) + \sum_{s \subseteq E_i} \lambda_s v_s(X) \end{aligned}$$

   as required.

**(Right-to-left)** To show the right-to-left direction, suppose that $\vec{c} = \langle c, a \rangle$ is in the convex hull of $\{\vec{c}_s : s \in \mathcal{S}\}$, i.e.,

$$\vec{c} = \sum_{s \in \mathcal{S}} \lambda_s \vec{c}_s$$

First, we note that $\lambda_s = c(s)$. By assumption, $c(s) = \sum_{s' \in \mathcal{S}} \lambda_{s'} v_{s'}(s)$. But $v_{s'}(s) = 1$, if $s = s'$; and $v_{s'}(s) = 0$, if $s \neq s'$. Thus, $c(s) = \lambda_s$, as required.

This allows us to infer the following: for each evidence proposition $E_i$ and proposition $X$,

$$\begin{aligned} a_{E_i}(X) &= \sum_{s \not\subseteq E_i} \lambda_s a_{E_i}(X) + \sum_{s \subseteq E_i} \lambda_s v_s(X) \\ &= \sum_{s \not\subseteq E_i} c(s)a_{E_i}(X) + \sum_{s \subseteq E_i} c(s)v_s(X) \\ &= (1 - c(E_i))a_{E_i}(X) + c(X \cap E_i) \end{aligned}$$

And from this, we obtain:

$$c(E_i)a_{E_i}(X) = c(X \cap E_i)$$

That is, $\vec{c} = \langle c, a \rangle$ is a conditionalizing strategy.

**Reformatting Inaccuracy Scores** In the accuracy-dominance argument for Probabilism, Predd et al. [2009] prove the following: if $\mathfrak{I}$ is a separable and continuous strictly proper inaccuracy measure, and $c$ is a non-probabilistic credence function, then there is a probabilistic credence function $c^*$ such that $\mathfrak{I}_s(c^*) < \mathfrak{I}_s(c)$ for all states $s$. A crucial component in that proof is the following idea: Intuitively, the inaccuracy of a credence function is its distance from the truth. More precisely, the inaccuracy of a credence function, $c$, at a state of the world, $s$, is the distance from the valuation function of that state (namely, $v_s$) to the credence function (namely, $c$). And indeed, as Predd et al. show, each separable and continuous strictly proper inaccuracy measure is generated as follows: take a certain sort of distance function; use that distance function to measure the distance from the valuation function $v_s$ to the credence function $c$; and take the inaccuracy of $c$ at $s$ to be given by that distance. The class of distance functions is the class of *additive Bregman divergences*.

Each additive Bregman divergence is characterised by a continuous, differentiable, strictly convex function $f : [0,1] \to [0,\infty)$ [Banerjee et al., 2005]. Given such an $f$, the corresponding additive Bregman divergence $\mathfrak{D}_f : [0,1]^n \times [0,1]^n \to [0,\infty]$ is

$$\mathfrak{D}_f(\vec{x},\vec{y}) = \sum_{i=1}^{n} f(x_i) - f(y_i) - f'(y_i)(x_i - y_i)$$

where $f'$ is the first derivative of $f$. One famous additive Bregman divergence is squared Euclidean distance: that is, $SED(\vec{x},\vec{y}) = \sum_i (x_i - y_i)^2$. A little calculation shows that $SED$ is generated by $f(x) = x^2$. Thus, the first crucial result from Predd, et al. is this:

**Lemma 1** $\mathfrak{I}$ is a separable and continuous strictly proper inaccuracy measure iff there is an additive Bregman divergence $\mathfrak{D}$ such that $\mathfrak{I}_s(c) = \mathfrak{D}(v_s, c)$ for all states $s$ and credence functions $c$.

This idea — that separable and continuous inaccuracy measures are generated by additive Bregman divergences — will be crucial in our proof as well. Thus, we prove the following claim:

**Lemma 2** If $\mathfrak{I}_s(c)$ is a proper scoring rule, then there is a Bregman divergence $\mathfrak{D}$ such that

$$\mathfrak{I}_s(\langle c, a\rangle) = \mathfrak{D}(\vec{c_s}, \vec{c})$$

Proof of Lemma 2: Suppose $\mathfrak{I}$ is a proper scoring rule. Then, by Lemma 1, there is a

Bregman divergence $\mathfrak{D}$ such that, for any credence function $c$, $\mathfrak{I}_s(c) = \mathfrak{D}(v_s, c)$. Now suppose $s$ is a state that entails $E_j$. Then

$$\begin{aligned}
\mathfrak{D}(\vec{c}_s, \vec{c}) &= \mathfrak{D}(v_s, c) + \mathfrak{D}(v_s, a_{E_j}) + \sum_{i \neq j} \mathfrak{D}(a_{E_i}, a_{E_i}) \\
&= \mathfrak{D}(v_s, c) + \mathfrak{D}(v_s, a_{E_j}) \text{ (since } \mathfrak{D}(a_{E_i}, a_{E_i}) = 0 \text{ for all } i) \\
&= \mathfrak{I}_s(c) + \mathfrak{I}_s(a_{E_j}) \\
&= \mathfrak{I}_s(\langle c, a \rangle) \text{ (by Temporal Separability)}
\end{aligned}$$

**Finding a Dominating Strategy** The accuracy-dominance argument that Predd et al. give is based on the following fact about Bregman divergences:

**Lemma 4** Let $\mathfrak{D}$ be an additive Bregman divergence. And let $\mathcal{X}$ be a set of vectors. Then if the vector $z$ lies outside the closed convex hull of $\mathcal{X}$, then there is another vector $z^*$ that lies in the convex hull of $\mathcal{X}$ such that $\mathfrak{D}(x, z^*) < \mathfrak{D}(x, z)$ for all $x$ in $\mathcal{X}$ — that is, for each member $x$ of $\mathcal{X}$, the divergence from $x$ to $z^*$ is less than the divergence from $x$ to $z$.

This fact will be crucial in our proof as well — but we won't prove it here.

Suppose $\vec{c} = \langle c, a \rangle$ is not a conditionalizing strategy. Then, as we have just established, $\vec{c}$ lies outside the closed convex hull of the $\vec{c}_s$s — since the set of $\vec{c}_s$s is finite, its convex hull is guaranteed to be closed. So, by the fact just stated, there is a vector $\vec{c^*}$ inside that convex hull that is closer to each $\vec{c}_s$ than $\vec{c}$ is. Furthermore, $\vec{c^*}$ corresponds to a unique strategy $\langle c^*, a^* \rangle$, where

$$\vec{c^*} = c^* \frown a^*_{E_1} \frown \ldots \frown a^*_{E_n}$$

Now, we know that, for each $\vec{c}_s$, $\mathfrak{D}(\vec{c}_s, \vec{c^*}) < \mathfrak{D}(\vec{c}_s, \vec{c})$. But, from Lemma 1 above, we know that $\mathfrak{D}(\vec{c}_s, \vec{c^*}) = \mathfrak{I}_s(\langle c^*_0, a^* \rangle)$ and $\mathfrak{D}(\vec{c}_s, \vec{c}) = \mathfrak{I}_s(\langle c, a \rangle)$. Thus, we know that, for each state $s$,

$$\mathfrak{I}_s(\langle c^*_0, a^*_{\mathcal{E}} \rangle) < \mathfrak{I}_s(\langle c, a \rangle)$$

That is, $\vec{c} = \langle c, a \rangle$ is strongly accuracy-dominated by $\vec{c^*} = \langle c^*, a^* \rangle$, as required. This completes the proof of (I') from above. Recall: (I') is the seemingly weaker version of (I). It claims only that each credal strategy that isn't both probabilistic and conditionalizing is dominated. It does not say anything about the sort of credal strategies that do the dominating. We now use (I') to establish (I).

**Finding a Probabilistic and Conditionalizing Dominating Strategy** Suppose that there are $k$ states. Thus, $\mathcal{S} = \{s_1, \ldots, s_k\}$. Then, for any credal strategy $\vec{c} = \langle c, a \rangle$, let

$$\mathfrak{I}(\vec{c}) := \langle \mathfrak{I}_{s_1}(\vec{c}), \ldots, \mathfrak{I}_{s_k}(\vec{c}) \rangle \in [0, \infty]^n$$

We will call this the *inaccuracy vector* of the credal strategy $\vec{c}$. Given two credal strategies $\vec{c}_1$ and $\vec{c}_2$, we write $\mathfrak{I}(\vec{c}_1) < \mathfrak{I}(\vec{c}_2)$ if $\mathfrak{I}_{s_i}(\vec{c}_1) < \mathfrak{I}_{s_i}(\vec{c}_2)$ for all $1 \leq i \leq n$. That is, $\mathfrak{I}(\vec{c}_1) < \mathfrak{I}(\vec{c}_2)$ if $\vec{c}_1$ strongly accuracy-dominates $\vec{c}_2$.

Suppose $\vec{c}_0, \ldots, \vec{c}_\alpha, \ldots$ is a transfinite sequence of credal strategies (where the sequence is defined on the ordinal $\lambda$). And suppose $\mathfrak{I}(\vec{c}_\beta) < \mathfrak{I}(\vec{c}_\alpha)$ for all $\beta > \alpha$ — that is, each credal strategy strongly accuracy-dominates all earlier ones. Then, since $\mathfrak{I}(\vec{c})$ is bounded below by $\langle 0, \ldots, 0 \rangle$, we have that the sequence $\mathfrak{I}(\vec{c}_0), \ldots, \mathfrak{I}(\vec{c}_\alpha), \ldots$ converges to a limit, by a transfinite version of the Monotone Convergence Theorem. Further, by a transfinite version of the Bolzano-Weierstrass Theorem, there is a transfinite subsequence $\vec{c}_{i_0}, \ldots, \vec{c}_{i_\alpha}, \ldots$, unbounded in the original sequence (and defined on ordinal $\gamma \leq \lambda$), that converges to a limit. Let that limit be $\vec{c}$. So $\lim_{\alpha < \gamma} \vec{c}_{i_\alpha} = \vec{c}$. Then

$$\lim_{\alpha < \lambda} \mathfrak{I}(\vec{c}_\alpha) = \lim_{\alpha < \gamma} \mathfrak{I}(\vec{c}_{i_\alpha}) = \mathfrak{I}(\vec{c})$$

Thus, $\vec{c}$ is a credal strategy whose inaccuracy vector is the limit of the inaccuracy vectors of the pairs in the original sequence. As a result, $\mathfrak{I}(\vec{c}) < \mathfrak{I}(\vec{c}_\alpha)$, for all $\alpha < \lambda$.

Suppose $\vec{c} = \langle c, a \rangle$ is a non-conditionalizing strategy. Then we can define the following sequence of credal strategies by transfinite recursion on the first uncountable ordinal.

- BASE CASE $\vec{c}_0 = \langle c, a \rangle$
- SUCCESSOR ORDINAL $\vec{c}_{\lambda+1}$ is any pair that strongly accuracy dominates $\vec{c}_\lambda$, if such exists; and $\vec{c}_\lambda$, if not.
- LIMIT ORDINAL $\vec{c}_\lambda$ is the strategy defined as above whose inaccuracy vector is the limit of the inaccuracy vectors of the strategies $\vec{c}_\alpha$ for $\alpha < \lambda$.

Then we can show that there must be $\alpha$ such that $\vec{c}_\alpha = \vec{c}_{\alpha+1}$. After all, there are at most continuum-many distinct pairs in the list $\mathfrak{I}(\vec{c}_0), \ldots, \mathfrak{I}(\vec{c}_\alpha), \ldots$. Thus, $\vec{c}_\alpha$ dominates the non-conditionalizing strategy $\vec{c}_0 = \langle c, a \rangle$. But $\vec{c}_\alpha$ is not itself dominated. Thus, $\vec{c}_\alpha$ must be a conditionalizing strategy, as required.

This completes the proof of (II). QED.

## Proof of Theorem 1(III)

Suppose that $\langle c, a \rangle$ is probabilistic and conditionalizing. Then, by Strict Propriety and the supposition that $c$ is a probability function, $c$ assigns itself a strictly higher degree of expected accuracy than it assigns to any other credal act. By Greaves and Wallace's result, $a$ enjoys maximal expected accuracy from the vantage point of $c$. So, by Temporal Separability, $\langle c, a \rangle$ enjoys minimal expected inaccuracy from the vantage point of $c$. But a strategy that weakly accuracy-dominated $\langle c, a \rangle$ would have expected inaccuracy greater than or equal to the expected inaccuracy of $\langle c, a \rangle$. Therefore, no other strategy weakly accuracy-dominates $\langle c, a \rangle$.

This completes the proof of (III). QED.