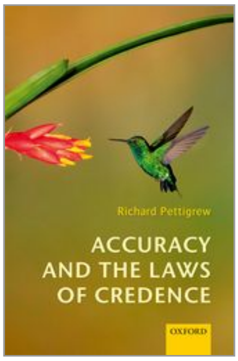
University Press Scholarship Online
Oxford Scholarship Online



Accuracy and the Laws of Credence
Richard Pettigrew

Print publication date: 2016
Print ISBN-13: 9780198732716
Published to Oxford Scholarship Online: May 2016
DOI: 10.1093/acprof:oso/9780198732716.001.0001

Diachronic Conditionalization

Richard Pettigrew

DOI:10.1093/acprof:oso/9780198732716.003.0016

Abstract and Keywords

This chapter asks whether there are any diachronic principles of credal rationality corresponding to the synchronic principles for updating rules that were considered in the previous chapter. The chapter concludes that there are no such principles.

Keywords: Diachronic rationality, diachronic continence, Conditionalization, evidence, update rules

So far, we have considered arguments for *planning* to update by conditionalizing; we have not considered any argument for actually updating by conditionalizing. In this chapter, we survey two attempts at such an argument and conclude that neither works. Lessons from the second suggest that no such argument can work.

15.1 The argument from the prior standpoint

We begin in this section with a strategy by which we might seek to establish Diachronic Conditionalization, which is the following diachronic requirement of rationality.[1]

Diachronic Conditionalization Suppose an agent has credence function c at t and c' at t' (with $t \leq t'$); and suppose that E is our agent's total evidence at t'. Then it is a requirement of rationality that, for all propositions X in $\mathscr{F}$,

$$c'(X) = c(X|E)$$

Indeed, we show that this strategy does give rise to an argument for that diachronic requirement when we measure inaccuracy in a particular way. However, we will see that the inaccuracy measures involved do not satisfy the conditions we introduced in Chapter 4—that is, they are not generated by additive Bregman divergences in conjunction with Alethic Vindication. Indeed, they cannot be used to give an accuracy dominance argument for Probabilism because it is not the case that every non-probabilistic function is accuracy dominated when they are used to measure accuracy. We then turn to consider the brute updating principles that are justified using this argument strategy. Finally, we question a further assumption of the argument.

Suppose that, at t, I have credence function c. Then I learn a proposition E with certainty. How does rationality require me to respond? This is the situation governed by Diachronic Conditionalization. How are we to justify that principle? Here is one suggestion. According to Greaves and Wallace's future-facing justification of Plan Conditionalization, at time t, prior to receiving evidence E, rationality (p.200) evaluates my updating plan by appealing to c, my credence function at that time, and considering what minimizes expected inaccuracy by its lights. According to the present argument, even after I have received the evidence E, rationality still evaluates my epistemic options by appealing to c and considering what minimizes expected inaccuracy by its lights. But in this case, it is not my updating plans that rationality is evaluating, since the evidence is now in and there is no longer any need for planning. Rather, rationality is appealing to c to evaluate possible posterior credence functions—that is, credence functions that I may adopt in response to the evidence E.

Now, there is a tension here: I must adopt a new posterior credence function because my current one—namely, c— doesn't respect my evidence. It assigns credence less than 1 to E. So I know that c is defective. If it weren't, I wouldn't need to adopt a replacement. Yet we assess the rationality of my possible posterior credence functions by appealing to that very credence function, the one we know to be defective. Is this ever warranted? I think perhaps it is. It seems that we might discover at least a rationally *permissible* posterior credence function by asking what my admittedly defective prior credence function c recommends, since there is no other credence function available whose recommendations have any greater legitimacy than the recommendations of c. Of course, there are plenty of credence functions around that respect the evidence in the sense that they assign credence 1 to E, so they are better than c in that respect. But the very question at issue is which of those credence functions rationality requires me to adopt. So, whatever c recommends is a rationally *permissible* posterior. But Diachronic Conditionalization requires more than that. It posits a rationally *mandated* posterior. It is much less obvious that, upon learning E and thus learning that c is defective, what is rationally *required* of me is whatever c recommends. Nonetheless, let us grant that. How, then, might I use c to choose my posterior credence function c' at t'?

The idea is this: First, my new credence function c' at t' must respect my evidence. That is, $c'(E) = 1$. Second, in line with Veritism, I wish to minimize inaccuracy. Remember, we are granting for the sake of argument that what rationality requires of me is what my defective credence function c recommends. Thus, rationality must require that I adopt the credence function amongst those that respect my evidence that minimizes expected inaccuracy by the lights of c. This is essentially the strategy employed in (Leitgeb & Pettigrew, 2010b).

Which credence function does minimize expected inaccuracy in this way? It turns out that it depends on the way in which I measure inaccuracy. Recall the logarithmic inaccuracy measure from Part III:

$$\mathfrak{L}(c, w) := -\ln c(w)$$

Then we have the following theorem, which is related to Theorem 5.1 of (Diaconis & Zabell, 1982) and noted in (Levinstein, 2012).

(p.201) **Theorem 15.1.1** *Suppose E is in $\mathcal{F}$. And suppose c is a probabilistic credence function with c(E) > 0. Then*

$$\lim_{\alpha \to 1}\left(\underset{\substack{p \in \mathcal{P} \\ p(E)=\alpha}}{\operatorname{argmin}} \operatorname{Exp}_{\mathfrak{L}}(p|c)\right) = c(-|E)$$

This theorem treats the problem that interests us by an indirect route. We're interested in which credence function amongst those that respect evidence E has minimal expected inaccuracy by the lights of c. A credence function respects evidence E if it assigns credence 1 to that evidence. However, we cannot simply try to minimize $\operatorname{Exp}_{\mathfrak{L}}(p|c)$ over those p with $p(E) = 1$. After all, all such credence functions assign $p(w) = 0$ for w not in E. But then $\mathfrak{L}(p, w) = -\ln p(w) = \infty$ for such w. So $\operatorname{Exp}_{\mathfrak{L}}(p|c) = \infty$ for all p (providing $c(E) < 1$). Thus, all such p minimize this quantity. So instead we take each $\alpha < 1$ and we ask which probabilistic credence function amongst those that assign credence α to E (and thus

$1 - \alpha$ to $\bar{E}$) minimizes expected inaccuracy by the lights of c. Once we have each of those, we then take their limit as α tends to 1 (if such exists). As the theorem tells us, it turns out to be $c(-|E)$, the posterior credence function mandated by Diachronic Conditionalization.

Thus, if an agent adopts $\mathfrak{L}$ as her inaccuracy measure, the argument would entail that she is rationally required to update in line with Diachronic Conditionalization. However, there is a problem: this inaccuracy measure does not satisfy the conditions on inaccuracy measures that we formulated in Chapter 4. As we noted in footnote 3, Chapter 12, it is strictly $\mathcal{P}$-proper but not strictly proper. Also, it is easy to see that it is not additive since it takes the inaccuracy of a credence function to be a function only of the values that credence function takes at the possible worlds—it pays no attention to the credences assigned to less specific propositions in the set, such as propositions that are true at two or more of the possible worlds. This feature prevents $\mathfrak{L}$ from providing an accuracy dominance argument for Probabilism. This is particularly worrying since the only way that we could show that the posterior that minimizes expected inaccuracy relative to $\mathfrak{L}$ amongst credence functions that respect the evidence is the one demanded by Diachronic Conditionalization was by restricting our attention to only the probabilistic credence functions. Without that restriction, there is no unique minimizer—indeed, any credence function that assigns credence 1 to every possible world and credence 1 to E will respect the evidence and will have maximal expected inaccuracy (namely, 0).

This raises the following question: Are there any inaccuracy measures that are legitimate by the lights of our characterization in Chapter 4—or that are not, but can at least provide an accuracy dominance argument for Probabilism—and that can be used to argue for Diachronic Conditionalization in the way mooted? I have no complete answer, but the following result does not inspire hope.

(p.202) First, some definitions:

Definition 15.1.2 (**Supervenes on difference**) *Suppose $\mathfrak{s}$ is a scoring rule. We say that $\mathfrak{s}$* supervenes on difference if $\mathfrak{s}(0, x) = \mathfrak{s}(1, 1 - x)$.

Definition 15.1.3 (**Continuously differentiable**) *Suppose $\mathfrak{s}$ is a scoring rule. We say that $\mathfrak{s}$* is continuously differentiable *if $\mathfrak{s}(0, x)$ and $\mathfrak{s}(1, x)$ are both differentiable as functions of x with continuous derivatives.*

Note that the quadratic, spherical, and logarithmic scoring rules are all continuously differentiable in this sense, and supervene on difference.

Theorem 15.1.4 *Suppose $\mathfrak{I}$ is generated by a continuously differentiable strictly proper scoring rule $\mathfrak{s}$ that supervenes on difference. Then there is a probabilistic credence function c defined on $\mathcal{F}$ and a proposition E in $\mathcal{F}$ such that*

$$\underset{\substack{p \in P \\ p(E)=1}}{\arg\min}\ \mathrm{Exp}_{I}(p|c) \neq c(-|E)$$

Thus, if you use an inaccuracy measure generated by a continuously differentiable strictly proper scoring rule that supervenes on difference —a scoring rule such as the quadratic, spherical, or logarithmic scoring rule, for instance—then there will be some proposition you might learn such that the probabilistic credence function amongst those that respect that evidence that minimizes expected inaccuracy relative to that inaccuracy measure will not be the one recommended by Diachronic Conditional- ization. So, if many of the familiar inaccuracy measures do not justify Diachronic Conditionalization, what updating rule do they recommend?

In the case of the Brier score, Hannes Leitgeb and I gave the answer in (Leitgeb & Pettigrew, 2010b).

Theorem 15.1.5 *Suppose c is a probabilistic credence function on $\mathcal{F}$ and E is a proposition in $\mathcal{F}$ with c(E) > 0. Then*

$$\underset{\substack{p \in P \\ p(E)=1}}{\arg\min}\ \mathrm{Exp}_{B}(p|c) = c(-\parallel E)$$

where

$$c(X \parallel E) = c(X \,\&\, E) + \frac{|X \,\&\, E|}{|E|}(1 - c(E))$$

Thus, if your credence function is c and you use the Brier score to measure inaccuracy—the inaccuracy measure generated by the quadratic scoring rule—then you should respond to evidence E by adopting credence function $c(-\parallel E)$, not $c(-|E)$, the credence function recommended by Diachronic Conditionalization. What can we say about this rival updating rule?[2] The joint paper with Hannes Leitgeb from which (p.203) some of the results in this chapter are drawn was spurred by Leitgeb's observation that this rule—the one that demands $c(-\parallel E)$ upon receipt of E—is in fact a particular member of the family of updating rules known as *imaging rules* that has been considered by Jim Joyce in the context of causal decision theory (Joyce, 2010b).

To see this, let's quickly recite the basics about imaging. The idea is to say what credence an agent should have in a proposition X *under the subjunctive supposition that E is true*, given her credences in the absence of that supposition. This is in contrast with the ratio definition of conditional probabilities, which is often taken to say what credence an agent should have in X *under the indicative supposition that E is true*, given her credences in the absence of that supposition. Different accounts of rational subjunctive supposition arise from different *transfer functions*. A transfer function $\rho E(w, w')$ takes a proposition E and two worlds w and w' and it determines the proportion of the probability that c assigns to w that should be transferred to w' in the event that we suppose E subjunctively. That is, if wewrite $c(X \| \rho E)$ for our credence in X on the subjunctive supposition of c relative to the transfer function ρ, we have

$$c(w' \|_\rho E) = \sum_{w \in W_F} c(w) \rho_E(w, w')$$

And then, of course,

$$c(X \|_\rho E) = \sum_{w' \in X} c(w' \|_\rho E)$$

So the credence that $c(- \|_\rho E)$ assigns to a world w' is obtained by working through each possible world w, taking the proportion of $c(w)$ that ρ demands we transfer to w', and summing these together. Thus, for instance, if $\rho e(w, w') = 1$, then, under the supposition of E, all of the probability that c assigns to w will be transferred to w'.

Here are three examples of transfer functions:

Example 15.1.6 *Suppose that, for each world w and each proposition E, there is a single closest E-world to w—call it* w^E*. Then we might define the following transfer function*:

$$\sigma_E(w, w') = \begin{cases} 1 & \text{if } w' = w^E \\ 0 & \text{if } w' \neq w^E \end{cases}$$

This is a Stalnakerian transfer function.

However, as is well known, there are serious problems with the assumption that, for every world and every proposition, there is a unique closest world at which the proposition is true. If there is then, according to the standard possible worlds semantics for subjunctive conditionals, the principle called Conditional Excluded Middle holds. This states that, for any two propositions A and B, one of the following two conditionals is true: *If A were the case, then B would be the case* or *If A were the case, then B would not be the case*. Thus, for instance, one of the following must be true: *If I toss the coin in my pocket, it will* *land heads or If I toss the coin in my pocket,* (p.204) *it will land tails*. But that seems false. In response to this, David Lewis characterized a broader class of transfer functions (Lewis, 1981):

> **Example 15.1.7** *Suppose that, for each world w and each proposition E, there is a set of closest E-worlds to w—call it w[E]. Then we might require the following of our transfer function*:
>
> (i) *If* $w' \notin w[E]$, *then* $\delta_e(w, w') = 0$
> (ii) *For each* w,
> $$\sum_{w' \in w[E]} \delta_E(w, w') = 1$$

This is a Lewisian transfer function.

This is quite a liberal definition. All that it requires is that no probability is transferred from any world to a world outside *w[E]*; it is all transferred into *w[E]*. Jim Joyce identified a particular Lewisian transfer function in the case where *w[E]* = *E* (Joyce, 2010b, 149).

> **Example 15.1.8**
>
> $$\lambda_E(w_i, w_j) = \begin{cases} 1 & \text{if } w_i, w_j \in E \text{ and } w_i = w_j \\ 0 & \text{if } w_i, w_j \in E \text{ and } w_i \neq w_j \\ \frac{1}{|E|} & \text{if } w_i \notin E \text{ and } w_j \in E \\ 0 & \text{if } w_j \notin E \end{cases}$$

We follow Joyce in calling this the Laplacian transfer function.

Let us see the imaging rule to which the Laplacian transfer function gives rise.

> (i) If $w \notin E$, then $c(w \|_\lambda E) = 0$.
> (A gloss: No credence should be assigned to possibilities that are incompatible with the agent's evidence. In this respect, this imaging rule agrees with conditionalization.)
> (ii) If $w \in E$, then
> $$\begin{aligned} c(w \|_\lambda E) &= \sum_{w' \in \mathrm{W_F}} c(w') \lambda_E(w', w) \\ &= c(w) + \frac{1 - c(E)}{|E|} \end{aligned}$$
> And thus
> $$c(X \|_\lambda E) = c(X \,\&\, E) + \frac{|X \,\&\, E|}{|E|}(1 - c(E))$$
> (A gloss: Suppose you learn E. Conditionalization suggests that you remove all of the credence that you assigned to the $\bar{E}$-worlds and redistribute it over (p.205) the E-worlds *in proportion to the credences that you previously assigned to those possibilities*. In

contrast, Brute Laplacian Imaging suggests that you remove all of the credence that you assigned to the Ē-worlds and redistribute it over the E-worlds, *adding the same amount to your previous credence in each such possibility*.)

Now, recall that $c(-\|_\lambda)$ is precisely the credence function that minimizes expected inaccuracy relative to the Brier score amongst all credence functions that assign maximal credence to E. That is,

$$\underset{\substack{p \in \mathcal{P} \\ p(E)=1}}{\operatorname{argmin}} \operatorname{Exp}_{\mathfrak{B}}(p|c) = c(-\|_\lambda E)$$

Thus, rather than giving us an argument for Diachronic Conditionalization, the Brier score—which, recall, is the only inaccuracy measure that satisfies all of the conditions laid out in Chapter 4, including Symmetry—gives an argument for Brute Laplacian Imaging instead.

Brute Laplacian Imaging Suppose an agent has credence function c at t and c' at t' (with $t \leq t'$); and suppose that E is our agent's total evidence at t'. Then it is a requirement of rationality that, for all propositions X in $\mathcal{F}$,

$$c'(X) = c(X \|_\lambda E)$$

This leaves us in a strange position: Suppose our agent knows that she will learn a proposition from the partition $\mathcal{E}$. Then, by any one of the arguments in Chapter 14, she is rationally required to plan to update by conditionalizing. Now suppose we fast forward to the moment at which she then obtains a piece of evidence E that belongs to $\mathcal{E}$. At that point, according to the argument that we have been considering here, the agent is rationally required to update by Laplacian Imaging—at least if their inaccuracy measure is the Brier score. What's more, they could know in advance that this would be the required way to update. So we have a situation in which an agent is rationally required to plan to do something that she knows she will be rationally required not to do when the time comes. It seems that something has gone wrong.

I think that what has gone wrong is this: When I evaluate the expected utility of a plan over a partition, I consider how good each outcome of the plan would be only at those worlds where it is the outcome of the plan. For instance, if I am evaluating the expected inaccuracy of an updating plan over a partition, I consider each element of the partition, I consider the response to that element that the updating plan endorses, and I look only at the inaccuracy ofthat response at the worlds at which the response would be made, weighting each by the probability I assign to that world. This isn't the case when we evaluate different possible posterior credence functions after we have obtained that evidence. In that case, we look at all the credence functions that respect this evidence. But we then consider not only how inaccurate they are at worlds at (p.206) which the evidence is true—weighting those inaccuracies by the probabilities our prior credence function assigns to those worlds—but also at worlds at which the evidence is false—again, applying the probabilistic weighting. It is this difference that gives rise to the difference in recommendation. If, in the second situation—what we might call the post hoc case—we were to restrict the set of worlds at which we consider the weighted inaccuracy of the posterior credence function, we obtain another argument for Conditionalization.

Theorem 15.1.9 *Suppose $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. Then*

$$\underset{p\in B}{\operatorname{argmin}} \sum_{w\in E} c(w) I(c', w) = c(-|E)$$

Theorem 15.1.9 provides the basis for an argument for Diachronic Conditionalization, but, as we noted above, only in the presence of the assumption that what is rationally required of an agent at time t', having obtained evidence E between t and t', but having not yet responded to it, is whatever is judged best by the lights of c. However, this assumption is too strong. There is nothing irrational about an agent who simply abandons c as a guide to her decisions upon receipt of E.

15.2 The argument from diachronic continence

A more straightforward argument for Diachronic Conditionalization, which introduces no new technical apparatus and in fact simply piggybacks on the arguments for Plan Conditionalization, is available if we accept the following principle, which we call the principle of diachronic continence, following Sarah Paul (Paul, 2014):

Diachronic Continence Suppose that, at t, an agent intends to ϕ at t'. And suppose that, between t and t', the reasons relevant to this intention don't change. Then, at t', the agent is rationally required to ϕ.

It is not implausible that, when an agent adopts an updating rule, this amounts to her adopting an intention. It is an intention to respond to evidence she receives in a particular way. Moreover, if all that changes between the time t at which she forms this intention and the time t' where the evidence arrives is that she acquires the evidence, then it seems reasonable to say that the reasons relevant to the intention don't change. If I have an intention to respond in a particular way to each of a number of exclusive and exhaustive possible future events, we cannot count the occurrence of one of those events as changing the reasons relevant to the intention. I have considered the possibility of that event and I have formed an intention about how I will respond to it; nothing in my reasons for forming that intention changes if the event actually occurs. Thus, if Diachronic Continence is true, it allows us to move from Plan Conditionalization to Diachronic Conditionalization.

(p.207) But is it true? Or, more importantly: is it true in sufficient generality to encompass the case we are considering, viz., the case in which an agent forms an intention about how she will respond to evidence she might receive? Of course, there are practical reasons not to abandon intentions at the last minute, just before you have to take the action. Paul (2014, 342) and Bratman (2012) both cite the following three: sticking with your intentions avoids the cognitive costs of reappraising your decisions; intentions are often formed in an environment in which you are susceptible to the temptation to satisfy your short-term ends at the expense of your long-term goals, and intentions are intended to help you avoid that temptation; having formed an intention, you will often have already carried out actions that change the world in such a way that it would make the cost of following through on the intention less than the cost of adopting a different course of action. Bratman also argues in favour of Diachronic Continence from the assumption that we must, as agents, desire self-governance. That is, we must wish to act from a unified and stable standpoint. He argues that satisfying Diachronic Coherence is a necessary condition for this. Paul demurs, arguing that, while a desire to self-govern in many decisions may indeed be a necessary condition on being an agent, a desire for perfect self-governance—that is, a desire to self-govern in *all* decisions—is not:

> We also care about things like existential spontaneity, losing control, rolling the dice and letting the world decide, and other more Romantic ideals. For an agent with these multifaceted values, a life that is perfectly self-governed would not in fact be successful relative to her varied concerns. (Paul, 2014, 345)

Nonetheless, Paul does think that there are reasons for an agent to be diachronically continent, even if those reasons don’t lift the principle of Diachronic Continence to the status of a principle of rationality. She holds that our ability to make a commitment and stick to it is essential for our success in pursuing our goals. And she holds that this ability is psychologically undermined every time we violate Diachronic Continence. Each time we abandon our intentions at the last minute, we become less reliable to ourselves, leading us to trust ourselves less in the future when such reliability is essential.

How do these considerations affect the instances of Diachronic Continence that would warrant the move from Plan Conditionalization to Brute Conditionalizaton? In those cases, it seems to me, Paul's concern about Bratman's requirement of perfect self-governance is even more compelling. It is not clear that we need have any desire to self-govern our epistemic lives in Bratman's sense. We do not aim to undertake our epistemic actions from a unified and stable standpoint. This is literally true because we don't undertake any epistemic actions at all—we don't choose our doxastic states. But it is also true that we don't value having doxastic states that are *as if* they were chosen from a unified and stable standpoint. Our sole aim in our epistemic life is accuracy—or so says Veritism, the guiding principle of this book. Unless we can show that unified and stable standpoints are more conducive to that goal, we cannot show (p.208) there is an epistemic reason to favour epistemic self-governance. Of course, there may be pragmatic reasons: frequently performing epistemic flip-flops, or engaging in epistemic 'brute shuffling' as Richard Kraut calls it, is likely to have many of the pragmatic disadvantages that pragmatic brute shuffling does. But that gives us no epistemic argument for Diachronic Conditionalization, only a pragmatic one.

I conclude that there is no epistemic argument from Plan Conditionalization to Diachronic Conditionalization via Diachronic Continence. Indeed, the foregoing considerations seem to suggest that there can be no epistemic argument for Diachronic Conditionalization at all. After all, Diachronic Conditionalization requires us to set our posterior credences on the basis of our prior credences—it requires us to base our judgements at t' on our judgements at t. But, as we saw above, there is no epistemic reason that compels us to retain at t' any faith in the judgements we made at t. Doing so does not serve the goal of accuracy. Thus, it seems that Plan Conditionalization is the strongest rational principle we can obtain in this vicinity on the basis of purely epistemic considerations.

(p.209) Appendix V: The mathematical results

In this chapter, we prove the mathematical results that underpin the arguments given in this part of the book.

V.A Proof of Theorem 14.1.1

Theorem 14.1.1 (**Greaves and Wallace**) Suppose $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$. So $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. Suppose c is a probabilistic credence function. And suppose $\mathcal{E}$ is a partition. Then an updating rule $R_{\mathcal{E}}$ minimizes expected inaccuracy by the lights of c iff $R_{\mathcal{E}}$ is a conditionalization rule on $\mathcal{E}$ for c.

That is,

(i) If $R_{\mathcal{E}}$ and R'_{E} are both conditionalization rules on $\mathcal{E}$ for c, then

$$\mathrm{Exp}_I(R_{\mathrm{E}}|c) = \mathrm{Exp}_I(R'_{\mathrm{E}}|c)$$

(ii) If $R_{\mathcal{E}}$ is a conditionalization rule on $\mathcal{E}$ for c and R'_{E} is not, then

$$\mathrm{Exp}_I(R_{\mathrm{E}}|c) < \mathrm{Exp}_I(R'_{\mathrm{E}}|c)$$

Proof.

(i) Suppose $R_{\mathcal{E}}$ and R'_{E} are both conditionalization rules on $\mathcal{E}$ for c. Now,

$$\mathrm{Exp}_I(R_{\mathrm{E}}|c) := \sum_{w \in \mathrm{W}} c(w) I(R_{\mathrm{E}}, w) = \sum_{w \in \mathrm{W}} c(w) I(R_{E_w}, w)$$

and

$$\mathrm{Exp}_I(R'_{\mathrm{E}}|c) := \sum_{w \in \mathrm{W}} c(w) I(R'_{\mathrm{E}}, w) = \sum_{w \in \mathrm{W}} c(w) I(R'_{E_w}, w)$$

But $R_{E_w} = R'_{E_w}$ for all E_w such that $c(E_w) > 0$. Now, if $c(w) > 0$, then $c(E_w) > 0$. So, if $c(w) > 0$, then $R_{E_w} = R'_{E_w}$ and thus $I(R_{\mathrm{E}}, w) = I(R'_{\mathrm{E}}, w)$. Therefore,

$$\mathrm{Exp}_I(R_{\mathrm{E}}|c) = \mathrm{Exp}_I(R'_{\mathrm{E}}|c)$$

as required.

(p.210) (ii) Suppose $R_{\mathcal{E}}$ is a conditionalization rule on $\mathcal{E}$ for c and R'_{E} is not. Then there is at least one E in $\mathcal{E}$ such that $c(E) > 0$ and $R'_E(-) \neq c(-|E) = R_E(-)$. Thus, since $\mathfrak{I}$ is strictly proper and R_E is a probability function, for any such E, we have

$$\mathrm{Exp}_I(R_E|R_E) < \mathrm{Exp}_I(R'_E|R_E)$$
$$\Longleftrightarrow$$
$$\sum_{w \in \mathrm{W}} R_E(w)\, I(R_E, w) < \sum_{w \in \mathrm{W}} R_E(w)\, I(R'_E, w)$$
$$\Longleftrightarrow$$
$$\sum_{w \in \mathrm{W}} c(w|E)\, I(R_E, w) < \sum_{w \in \mathrm{W}} c(w|E)\, I(R'_E, w)$$
$$\Longleftrightarrow$$
$$\sum_{w \in E} \frac{c(w)}{c(E)}\, I(R_E, w) < \sum_{w \in E} \frac{c(w)}{c(E)}\, I(R'_E, w)$$
$$\Longleftrightarrow$$
$$\sum_{w \in E} c(w)\, I(R_E, w) < \sum_{w \in E} c(w)\, I(R'_E, w)$$

On the other hand, for those E in $\mathcal{E}$ such that $c(E) > 0$ and $R_E = R'_E$, clearly we have

$$\sum_{w \in E} c(w)\, I(R_E, w) = \sum_{w \in E} c(w)\, I(R'_E, w)$$

And, if E in $\mathcal{E}$ is such that $c(E) = 0$, then

$$\sum_{w \in E} c(w)\, I(R_E, w) = 0 = \sum_{w \in E} c(w)\, I(R'_E, w)$$

Putting these together, this gives:

$$\sum_{E \in \mathrm{E}} \sum_{w \in E} c(w)\, I(R_E, w) < \sum_{E \in \mathrm{E}} \sum_{w \in E} c(w)\, I(R'_E, w)$$
$$\Longleftrightarrow$$
$$\sum_{w \in \mathrm{W}} c(w)\, I(R_{\mathrm{E}}, w) < \sum_{w \in \mathrm{W}} c(w)\, I(R'_{\mathrm{E}}, w)$$
$$\Longleftrightarrow$$
$$\mathrm{Exp}_I(R_{\mathrm{E}}|c) < \mathrm{Exp}_I(R'_{\mathrm{E}}|c)$$

as required.

This completes the proof of the theorem. □

(p.211) V.B Proof of Theorem 14.2.1

Theorem 14.2.1 (**van Fraassen**) Plan Conditionalization ⇔ Generalized Reflection Principle.

Proof. ⇒. Suppose $R_{\mathcal{E}}$ is a conditionalization plan on $\mathcal{E}$ for c. So $c(XE) = c(E)R_E(X)$ for all X. Then, for all X,

$$c(X) = \sum_{E \in \mathrm{E}} c(XE) = \sum_{E \in \mathrm{E}} c(E) R_E(X)$$

so c is a mixture of the R_Es.

⇐. Suppose $R_{\mathcal{E}}$ is an updating rule and c is a mixture of the R_Es. Then, there are $\lambda_E \geq 0$ with $\sum_{E \in \mathcal{E}} \lambda_E = 1$ such that, for all X

$$c(X) = \sum_{E \in \mathrm{E}} \lambda_E R_E(X)$$

Thus, in particular, for E' in $\mathcal{E}$, we have

$$c(E') = \sum_{E \in \mathrm{E}} \lambda_E R_E(E')$$

But, since $R_E(E') = 1$ if $E = E'$ and $R_E(E') = 0$ if $E \neq E'$, we have that $c(E') = \lambda_{E'}$. Thus,

$$c(X) = \sum_{E \in \mathrm{E}} c(E) R_E(X)$$

And from this, we get

$$c(XE) = c(E) R_E(X)$$

as required. □

V.C Proof of Theorem 14.3.1

Theorem 14.3.1 Suppose $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$. So $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure.

(I) If $\langle c, R_{\mathscr{E}} \rangle$ is not a conditionalizing pair, then there is a conditionalizing pair $\langle c^*, R^*_{\mathrm{E}} \rangle$ such that, for all $w \in \mathscr{W}_{\mathscr{F}}$,

$$I(\langle c^*, R^*_{\mathrm{E}} \rangle, w) < I(\langle c, R_{\mathrm{E}} \rangle, w)$$

(II) If $\langle c, R_{\mathscr{E}} \rangle$ is a conditionalizing pair, then for any pair $\langle c^*, R^*_{\mathrm{E}} \rangle$ such that $c \neq c^*$ or $R_E(-) \neq R^*_E(-)$ for some $c(E) > 0$,

$$\mathrm{Exp}_I(\langle c, R_{\mathrm{E}} \rangle | c) < \mathrm{Exp}_I(\langle c^*, R^*_{\mathrm{E}} \rangle) | c$$

(p.212) *Proof*. Since $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$, there is a one-dimensional Bregman divergence ∂ such that $\mathfrak{D}(c, c') = \sum_{X \in \mathscr{F}} \mathfrak{d}(c(X), c'(X))$ and so $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c) = \sum_{X \in \mathscr{F}} \mathfrak{d}(v_w(X), c(X))$ and thus

$$\begin{aligned} I(\langle c, R_{\mathrm{E}} \rangle, w) &= \mathrm{D}(v_w, c) + \mathrm{D}(v_w, R_{E_w}) \\ &= \sum_{X \in \mathrm{F}} \mathrm{d}(v_w(X), c(X)) + \mathrm{d}(v_w(X), R_{E_w}(X)) \end{aligned}$$

The first step of the proof is to characterize the conditionalizing pairs. Given a pair $\langle c, R_{\mathscr{E}} \rangle$, represent it as a vector as follows:

$$\langle c(X_1), \ldots, c(X_n), R_{E_1}(X_1), \ldots, R_{E_1}(X_n), \ldots, R_{E_m}(X_1), \ldots, R_{E_m}(X_n) \rangle$$

Another way to write this is as the concatenation of vectors:

$$c \frown R_{E_1} \frown R_{E_2} \frown \ldots R_{E_{j-1}} \frown R_{E_j} \frown R_{E_{j+1}} \frown \ldots \frown R_{E_m}$$

Now, given a world w in $\mathscr{W}_{\mathscr{F}}$ such that E_j is true at w and a pair $\langle c, R_{\mathscr{E}} \rangle$, define the following vector:

$$\begin{aligned} \langle c, R_{\mathrm{E}} \rangle_w := \langle & v_w(X_1), \ldots, v_w(X_n), R_{E_1}(X_1), \ldots, R_{E_1}(X_n), \ldots, \\ & R_{E_{j-1}}(X_1), \ldots, R_{E_{j-1}}(X_n), v_w(X_1), \ldots, v_w(X_n), R_{E_{j+1}}(X_1), \ldots, R_{E_{j+1}}(X_n), \\ & \ldots, R_{E_m}(X_1), \ldots R_{E_m}(X_n) \rangle \end{aligned}$$

Another way to write this is as the concatenation of vectors:

$$\langle c, R_{\mathrm{E}} \rangle_w := v_w \frown R_{E_1} \frown R_{E_2} \frown \ldots \frown R_{E_{j-1}} \frown v_w \frown R_{E_{j+1}} \frown \ldots \frown R_{E_m}$$

Lemma V.C.1 *$\langle c, R_{\mathscr{E}} \rangle$ is a conditionalizing pair iff its Vector representation is in the convex hull of* $\{\langle c, R_{\mathscr{E}} \rangle_w : w \in \mathscr{W}_{\mathscr{F}}\}$

Proof of Lemma V.C.1 (⇒) Suppose $\langle c, R_{\mathscr{E}} \rangle$ is a conditionalizing pair. Then

$$R_{E_i}(X)c(E_i) = c(XE_i)$$

We can then show that

$$\langle c, R_{\mathrm{E}} \rangle = \sum_{w \in \mathrm{W_F}} \lambda_w \langle c, R_{\mathrm{E}} \rangle_w$$

where $\lambda_w = c(w)$, since

- First:

$$c(X) = \sum_{w \in \mathrm{W_F}} c(w) v_w(X)$$

(p.213) • Second, since $\langle c, R_{\mathscr{E}} \rangle$ is a conditionalizing pair, $c(E_i)R_{E_i}(X) - c(XE_i) = 0$. So

$$\begin{aligned} R_{E_i}(X) &= R_{E_i}(X) - (c(E_i)R_{E_i}(X) - c(XE_i)) \\ &= (1 - c(E_i))R_{E_i}(X) + c(XE_i) \\ &= \sum_{w \notin E_i} c(w)R_{E_i}(X) + \sum_{w \in E_i} c(w)v_w(X) \end{aligned}$$

(⇐) Suppose

$$\langle c, R_{\mathrm{E}} \rangle = \sum_{w \in \mathrm{W_F}} \lambda_w \langle c, R_{\mathrm{E}} \rangle_w$$

Then it must be that $\lambda_w = c(w)$. And thus, by a similar equation to above $R_{E_i}(X)c(E_i) = c(XE_i)$, as required. □

Proof of Theorem 14.3.1(I) Suppose $\langle c, R_{\mathscr{E}} \rangle$ is not a conditionalizing pair. Thus, $\langle c, R_{\mathscr{E}} \rangle$ sits outside the convex hull of $\{\langle c, R_{\mathscr{E}} \rangle_w : w \in \mathscr{W}_{\mathscr{F}}\}$. Now, since the convex hull of $\{\langle c, R_{\mathscr{E}} \rangle_w : w \in \mathscr{W}_{\mathscr{F}}\}$ is a closed convex set and $\mathfrak{D}$ is an additive Bregman divergence, there is a closest point on the convex hull of $\{\langle c, R_{\mathscr{E}} \rangle w : w \in \mathscr{W}_{\mathscr{F}}\}$ to $\langle c, R_{\mathscr{E}} \rangle$. Let's call it $\langle c^*, R^*_{\mathrm{E}} \rangle$. And we have, for all $w \in \mathscr{W}_{\mathscr{F}}$:

$$\mathrm{D}(\langle c, R_{\mathrm{E}} \rangle_w, \langle c^*, R^*_{\mathrm{E}} \rangle) < \mathrm{D}(\langle c, R_{\mathrm{E}} \rangle_w, \langle c, R_{\mathrm{E}} \rangle)$$

But, if E_j is true at w,

$$\begin{aligned} &\mathrm{D}(\langle c, R_{\mathrm{E}} \rangle_w, \langle c, R_{\mathrm{E}} \rangle) = \\ &\sum_{i=1}^n \mathrm{d}(v_w(X_i), c(X_i)) + \\ &\sum_{i=1}^n \mathrm{d}(R_{E_1}(X_i), R_{E_1}(X_i)) + \ldots + \\ &\sum_{i=1}^n \mathrm{d}(R_{E_{j-1}}(X_i), R_{E_{j-1}}(X_i)) + \\ &\sum_{i=1}^n \mathrm{d}(v_w(X_i), R_{E_j}(X_i)) + \\ &\sum_{i=1}^n \mathrm{d}(R_{E_{j+1}}(X_i), R_{E_{j+1}}(X_i)) + \ldots + \\ &\sum_{i=1}^n \mathrm{d}(R_{E_m}(X_i), R_{E_m}(X_i)) \end{aligned}$$

And, since $\mathfrak{d}(R_{Ek}(X_i), R_{Ek}(X_i)) = 0$, that gives

$$\begin{aligned} \mathrm{D}(\langle c, R_{\mathrm{E}}\rangle_w, \langle c, R_{\mathrm{E}}\rangle) &= \sum_{i=1}^{n} \mathrm{d}(v_w(X_i), c(X_i)) + \sum_{i=1}^{n} \mathrm{d}(v_w(X_i), R_{E_j}(X_i)) \\ &= \sum_{i=1}^{n} \mathrm{s}(v_w(X_i), c(X_i)) + \sum_{i-1}^{n} \mathrm{s}(v_w(X_i), R_{E_j}(X_i)) \\ &= I(\langle c, R_{\mathrm{E}}\rangle, w) \end{aligned}$$

(p.214) Moreover, if E_j is true at w,

$$\begin{aligned} &\mathrm{D}(\langle c, R_{\mathrm{E}}\rangle_w, \langle c^*, R^*_{\mathrm{E}}\rangle) = \\ &\sum_{i=1}^{n} \mathrm{d}(v_w(X_i), c^*(X_i)) + \\ &\sum_{i=1}^{n} \mathrm{d}(R_{E_1}(X_i), R^*_{E_1}(X_i)) + \ldots + \\ &\sum_{i=1}^{n} \mathrm{d}(R_{E_{j-1}}(X_i), R^*_{E_{j-1}}(X_i)) + \\ &\sum_{i=1}^{n} \mathrm{d}(v_w(X_i), R^*_{E_j}(X_i)) + \\ &\sum_{i=1}^{n} \mathrm{d}(R_{E_{j+1}}(X_i), c_{E^*_{j+i}}(X_i)) + \ldots + \\ &\sum_{i=1}^{n} \mathrm{d}(R_{E_m}(X_i), c^*_{E_m}(X_i)) \end{aligned}$$

And, since $\mathrm{d}(R_{E_k}(X_i), R^*_{E_k}(X_i)) \geq 0$, this gives

$$\begin{aligned} \mathrm{D}(\langle c, R_{\mathrm{E}}\rangle_w, \langle c^*, R^*_{\mathrm{E}}\rangle) &\geq \sum_{i=1}^{n} \mathrm{d}(v_w(X_i), c^*(X_i)) + \sum_{i=1}^{n} \mathrm{d}(v_w(X_i), R^*_{E_j}(X_i)) \\ &= \sum_{i=1}^{n} \mathrm{s}(v_w(X_i), c^*(X_i)) + \sum_{i-1}^{n} \mathrm{s}(v_w(X_i), R^*_{E_j}(X_i)) \\ &= I(\langle c^*, R^*_{\mathrm{E}}\rangle, w) \end{aligned}$$

Thus, from

$$\mathrm{D}(\langle c, R_{\mathrm{E}}\rangle_w, \langle c^*, R^*_{\mathrm{E}}\rangle) < \mathrm{D}(\langle c, R_{\mathrm{E}}\rangle_w, \langle c, R_{\mathrm{E}}\rangle)$$

we can infer

$$I(\langle c^*, R^*_{\mathrm{E}}\rangle, w) < I(\langle c, R_{\mathrm{E}}\rangle, w)$$

This establishes that, for every pair $\langle c, R_{\mathscr{E}}\rangle$ that isn't a conditionalizing pair, there is another pair $\langle c^*, R^*_{\mathscr{E}}\rangle$ that dominates $\langle c, R_{\mathscr{E}}\rangle$. However, for all we have said, there is no reason to think that $\langle c^*, R^*_{\mathrm{E}}\rangle$ is itself a conditionalizing pair. We now show that there is always a conditionalizing pair $\langle c^*, R^*_{\mathrm{E}}\rangle$ that dominates $\langle c, R_{\mathscr{E}}\rangle$.

Suppose $\mathscr{W} = \{w_1, \ldots, w_n\}$. And for any pair $P = \langle c, R\rangle$, let

$$I(P) := \langle I(P, w_1), \ldots I(P, w_n)\rangle \in [0, \infty]^n$$

We call this the inaccuracy vector of P. We write

$$I(P) < I(P') \text{ iff } I(P, w_i) < I(P', w_i)$$

for all $1 \leq i \leq n$.

Suppose $P_0, \ldots, P_\alpha, \ldots$ is a transfinite sequence of pairs (where the sequence is defined on ordinal λ). And suppose $\mathfrak{I}(P_\beta) < \mathfrak{I}(P_\alpha)$ for all $\beta > \alpha$—that is, each pair dominates all earlier ones. Then, since $\mathfrak{I}(P)$ is bounded below by $\langle 0, \ldots, 0\rangle$, we have (p.215) that the sequence $\mathfrak{I}(P_0), \ldots, \mathfrak{I}(P_\alpha), \ldots$ 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 $P_{i_0}, \ldots, P_{i_\alpha}, \ldots$, , unbounded in the original sequence (and defined on ordinal $\gamma \leq \lambda$) that converges to a limit. Let that limit be P. So $\lim_{\alpha<\gamma} P_{i_\alpha} = P$. Then

$$\lim_{\alpha<\lambda} I(P_\alpha) = \lim_{\alpha<\gamma} I(P_{i_\alpha}) = I(P)$$

Thus, P is a pair whose inaccuracy vector is the limit of the inaccuracy vectors of the pairs in the original sequence. As a result, $\mathfrak{I}(P) < \mathfrak{I}(P_\alpha)$, for all $\alpha < \lambda$.

Suppose $\langle c, R\rangle$ is a non-conditionalizing pair. Then define the following sequence of pairs by transfinite recursion on the first uncountable ordinal.

- BASE CASE $P_0 = \langle c, R\rangle$
- SUCCESSOR ORDINAL $P_{\lambda+1}$ is any pair that accuracy dominates $P\lambda$, if such exists; and P_λ, if not.
- LIMIT ORDINAL P_λ is the pair P defined as above whose inaccuracy vector is the limit of the inaccuracy vectors of the pairs P_α for $\alpha < \lambda$.

Then we can show that there must be a such that $P_\alpha = P_{\lambda+1}$, since there are at most continuum-many distinct pairs in the list $\mathfrak{I}(P_0), \ldots, \mathfrak{I}(P_\alpha), \ldots$. Thus, P_α dominates the non-conditionalizing pair $P_0 = \langle c, R\rangle$. But P_α is not itself dominated. Thus, P_α must be a conditionalizing pair, as required. This completes the proof of (I). □

Proof of Theorem 14.3.1(II) Suppose $\langle c, R_\mathscr{E}\rangle$ is a conditionalizing pair. And suppose $\langle c, R_\mathscr{E}\rangle \neq \langle c^*, R^*_\mathscr{E}\rangle$. Then:

- Since $\mathfrak{I}$ is strictly proper, $\mathrm{Exp}_{;\mathfrak{I}}(c|c) \leq \mathrm{Exp}_{;\mathfrak{I}}(c^*|c)$ with equality iff $c = c^*$.
- By Theorem 14.1.1, $\mathrm{Exp}_I(R_\mathrm{E}|c) \leq \mathrm{Exp}_I(R^*_\mathrm{E}|c)$, with equality iff R^*_E is also a conditionalization rule on $\mathscr{E}$ for c (and thus $R_E(-) = R^*_E(-)$ for all E such that $c(E) > 0$).

Thus,

$$\mathrm{Exp}_I(\langle c, R_E \rangle | c) \leq I(\langle c^*, R^*_E \rangle | c)$$

with equality iff $c = c^*$ and $R_E(-) = R^*_E(-)$ for all E such that $c(E) > 0$. □

V.D Proof of Theorem 15.1.1

Theorem 15.1.1 Suppose E is in $\mathscr{F}$. And suppose c is a probabilistic credence function with $c(E) > 0$. then

$$\lim_{\alpha \to 1} \left(\underset{\substack{p \in \mathrm{P} \\ p(E) = \alpha}}{\mathrm{argmin}} \ \mathrm{Exp}_L(p|c) \right) = c(-|E)$$

(p.216) To solve our minimization problem, we use the KKT conditions:

Theorem V.D.1 *Suppose $f, g_1, \ldots, g_m, h_1, \ldots, h_n : \mathbb{R}^k \to \mathbb{R}$ are smooth functions. Consider the following minimization problem. Minimize $f(x_1, \ldots, x_k)$ relative to the following constraints*:

$$\begin{aligned} g_i(x_1, \ldots, x_k) &\leq 0 \ \ for\ i = 1, \ldots, m, \\ h_j(x_1, \ldots, x_k) &= 0 \ for\ j = 1, \ldots, n. \end{aligned}$$

If $\mathbf{x}^* = (x^*_1, \ldots, x^*_k)$ is a (non-singular) solution to this minimization problem, then there exist $\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i \ \ \nabla_{gi}(\mathbf{x}^*) \ \ + \sum_{j=1}^{n} \lambda_j \nabla h_j(\mathbf{x}^*) &= 0 \\ \mu_{ig_i}(\mathbf{x}^*) &= 0 \quad for\ i = 1, \ldots m \\ \mu_i &\geq 0 \quad for\ i = 1, \ldots, m \\ g_i(\mathbf{x}^*) &\leq 0 \quad for\ i = 1, \ldots, m \\ h_i(\mathbf{x}^*) &= 0 \end{aligned}$$

If, furthermore, f and g are convex functions, then the existence of $\mu_1, \ldots, \mu_m, \lambda_1, \ldots, \lambda_n$ e $\mathbb{R}$ *is sufficient for a solution to the minimization problem. If f is strictly convex, then their existence is sufficient for a unique solution.*

Corollary V.D.2 *The following two statements are equivalent, where c is a probabilistic credence function defined over a partition $\{w_1, \ldots, w_n\}$ and E is the proposition true at worlds $w_1, \ldots, w_k$* with $c(E) > 0$:

(i)

$$\underset{\substack{p \in \mathrm{P} \\ p(E) = 1}}{\mathrm{argmin}} \ \mathrm{Exp}_I(p|c) = \mathbf{x}^*$$

(ii) The following three conditions hold:

- $\sum_{i=1}^{k} x^*_1 = 1$
- $x^*_i = 0$ *for* $i = k + 1, \ldots, n$

- There are $\mu_1, \ldots, \mu_k \geq 0$ *and* $\lambda, \lambda_{k+1}, \ldots, \lambda_n$ *such that*
  - *For all* $i = 1, \ldots, k$,
  $$\partial \Big/ \partial x_i \mathrm{Exp}_I(x^*|c) - \mu_i + \lambda := 0$$
  - *For all* $i = k + 1, \ldots, n$,
  $$\partial \Big/ \partial x_i \mathrm{Exp}_I(x^*|c) + \lambda_i = 0$$

(p.217) *where we represent a probability function p over* $\{w_1, \ldots, w_n\}$ *as the vector* $\langle p_1, \ldots, p_n \rangle$, *where* $p_i = p(w_i)$.

Proof. Let

- $f(\mathbf{x}) = \mathrm{Exp}_{\mathfrak{I}}(\mathbf{x}|c)$
- $g_i(\mathbf{x}) = -x_i$
- $h(\mathbf{x}) = x_i + \ldots + x_k - 1$
- $h_i(\mathbf{x}) = x_i$, for $i = k + 1, \ldots, n$.

Proof of Theorem 15.1.1. We prove this by showing that

$$\underset{\substack{p \in \mathrm{P} \\ p(E)=\alpha}}{\mathrm{argmin}}\ \mathrm{Exp}_{\mathrm{L}}(p|c) = \alpha c(-|X) + (1-\alpha)c(-|\neg X)$$

Let

- $f(\mathbf{x}) = \mathrm{Exp}_{\mathfrak{I}}(\mathbf{x}|c)$
- $g_i(\mathbf{x}) = -x_{\mathrm{i}}$
- $h_1(\mathbf{x}) = x_i + \ldots + x_k - \alpha$
- $h_2(\mathbf{x}) = x_k + 1 + \ldots + x_n - (1 - \alpha)$

Then:

- For all $i = \mathrm{i}, \ldots, n$, let $\mu_i := 0$
- Let $\lambda_1 := -\frac{c(E)}{\alpha}$
- Let $\lambda_2 = :- \frac{1-c(E)}{1-\alpha}$

Now, from above, we have:

$$\partial \Big/ \partial x_i \mathrm{Exp}_{\mathrm{L}}(x|c) = -\frac{c_i}{X_i}$$

So, let $\mathbf{x}^* = \alpha c(-|E) + (1 - \alpha)c(-|\neg \mathrm{E})$. Thus,

- If $i = 1, \ldots,$ then $x_i^* = \alpha \frac{c_i}{c(E)}$
- If $i = k + 1, \ldots, n$, then $x_i^* = (1-\alpha)\frac{c_i}{1-c(E1)}$

So

- $i = 1, \ldots, k$, then

$$-\frac{c_i}{x_i^*} - \mu_i + \lambda_1 = \frac{c(X)}{\alpha} - 0 - \frac{c(X)}{\alpha} = 0$$

- $k + 1 = 1, \ldots, n$, then

$$-\frac{c_i}{x_i^*} - \mu_i + \lambda_2 = \frac{1 - c(X)}{1 - \alpha} - 0 - \frac{1 - c(X)}{1 - \alpha} = 0$$

as required. □

(p.218) V.E Proof of Theorem 15.1.4

Theorem 15.1.4 Suppose $\mathfrak{I}$ is generated by a continuously differentiable strictly proper scoring rule $\mathfrak{s}$ that supervenes on difference. Then there is a probabilistic credence function c defined on $\mathcal{F}$ and a proposition E in $\mathcal{F}$ such that

$$\underset{\substack{p \in \mathrm{P} \\ p(E)=1}}{\mathrm{argmin}} \mathrm{Exp}_I(p|c) \neq c(-|E)$$

Proof. Let $\mathcal{F} = \{w_1, w_2, w_3\}$ and suppose w_1, w_2, w_3 form a partition. Now suppose our agent learns $E = w_2 \vee w_3$. Then her posterior p must be such that $p(E) = 1$ and $p(\overline{E}) = 0$. Thus, here are the priors and the family of possible posteriors p_x that satisfy these constraints:

	w_1	w_2	w_3
c	α	β	γ
px	0	x	$1-x$

Then define the following function, which is the one we're looking to minimize:

$$\begin{aligned} f(x) := \mathrm{Exp}_I(p_x|c) &= \alpha I(p_x, w_1) + \beta I(p_x, w_2) + \gamma I(p_x, w_3) \\ &= \alpha[\mathfrak{s}(1, 0) + \mathfrak{s}(0, x) + \mathfrak{s}(0, 1 - x)] + \\ &\quad \beta[\mathfrak{s}(0, 0) + \mathfrak{s}(1, x) + \mathfrak{s}(0, 1 - x)] + \\ &\quad \gamma[\mathfrak{s}(0, 0) + \mathfrak{s}(0, x) + \mathfrak{s}(1, 1 - x)] \end{aligned}$$

Then

$$f'(x) = \alpha\mathfrak{s}'(0, x) - \alpha\mathfrak{s}'(0, 1 - x) + \beta\mathfrak{s}'(1, x) - \beta\mathfrak{s}'(0, 1 - x) + \gamma\mathfrak{s}'(0, x) - \gamma\mathfrak{s}'(1, 1 - x)$$

Now, suppose $\mathrm{Exp}(p_x|c)$ is uniquely minimized at $p_x(-) = c(-|E)$. Then

$$f'\left(\frac{\beta}{1-\alpha}\right) = 0$$

since $x = p_x(w_2) = c(w_2|w_2 \vee w_3) = \frac{\beta}{1-\alpha}$.

Thus,

$$f'\left(\frac{\beta}{1-\alpha}\right) = \alpha\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) - \alpha\mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right) + \beta\mathfrak{s}'\left(1, \frac{\beta}{1-\alpha}\right) - \beta\mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right) + \gamma\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) - \gamma\mathfrak{s}'\left(1, \frac{\gamma}{1-\beta}\right) = 0$$

Now, since $\mathfrak{s}$ is strictly proper,

$$\frac{\beta}{1-\alpha}\mathfrak{s}'\left(1, \frac{\beta}{1-\alpha}\right) + \frac{\gamma}{1-\alpha}\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) = 0$$

(p.219) and

$$\frac{\gamma}{1-\alpha}\mathfrak{s}'\left(1, \frac{\gamma}{1-\alpha}\right) + \frac{\beta}{1-\alpha}\mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right) = 0$$

And thus

$$\beta\mathfrak{s}'\left(1, \frac{\beta}{1-\alpha}\right) + \gamma\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) = 0$$

and

$$\gamma\mathfrak{s}'\left(1, \frac{\gamma}{1-\alpha}\right) + \beta\mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right) = 0$$

Thus,

$$f'\left(\frac{\beta}{1-\alpha}\right) = \alpha\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) - \alpha\mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right) = 0$$

which gives

$$\mathfrak{s}'\left(0, \frac{\beta}{1-\alpha}\right) = \mathfrak{s}'\left(0, \frac{\gamma}{1-\alpha}\right)$$

Thus, since we imposed no conditions on α, β, γ, it follows that, for any $0 \leq x \leq 1$,

$$\mathfrak{s}'(0, x) = \mathfrak{s}'(0, 1-x)$$

Now suppose $\mathfrak{s}$ supervenes on difference. Then $\mathfrak{s}(0, x) = \mathfrak{s}(1, 1-x)$. Thus,

$$\mathfrak{s}'(0, x) = -\mathfrak{s}'(1, 1-x)$$

Putting these together, we get $-\mathfrak{s}'(0, x) = \mathfrak{s}'(1, x)$. Now, since $\mathfrak{s}$ is strictly proper,

$$0 = x\mathfrak{s}'(1, x) + (1-x)\mathfrak{s}'(0, x) = (2x-1)\mathfrak{s}'(0, x)$$

for any $0 \leq x \leq 1$. Thus, if $\mathfrak{s}$ is continuously differentiable, then for any $0 \leq x \leq 1$

$$\mathfrak{s}'(0, x) = 0$$

Thus, $\mathfrak{s}(0, x) = mx + k$ for some constants m and k. And thus $\mathfrak{s}(1, x) = m(1-x) + k$. But this scoring rule is not strictly proper. This completes the proof. □

V.F Proof of Theorem 15.1.5

Theorem 15.1.5 Suppose c is a probabilistic credence function defined on the partition $\{w_1, \ldots, w_n\}$ and E is the proposition true at $w_1, \ldots, w_k$ and $c(E) > 0$. Then

$$\underset{\substack{p \in P \\ p(E)=1}}{\text{argmin}}\ \text{Exp}_B(p|c) = c(-\,\|\,E)$$

(p.220) where

$$c(X \,\|\, E) = c(X \,\&\, E) + \frac{|X\&E|}{|E|}(1 - c(E))$$

Proof. Recall Corollary V.D.2.

- If $i = 1, \ldots, k$, let $\mu_i := 0$
- If $i = k + 1, \ldots, n$, let $\lambda_i := 2c_i$
- Let $\lambda := -2\frac{1-c(E)}{|E|}$

Now, from above, we have:

$$\partial / \partial x_i \text{Exp}_B(\mathbf{x}|c) = -2c_i + 2x_i$$

So, let $\mathbf{x}^* = c(-\|E)$. So $x_i^* = c_i + \frac{1-c(E)}{|E|}$. Then

- If $i = 1, \ldots, k$ then
$$-2c_i + 2x_i^* - \mu_i + \lambda = -2c_i + 2\left(c_i + \frac{1-c(E)}{|E|}\right) - 0 - 2\frac{1-c(E)}{|E|} = 0$$
- If $i = k + 1, \ldots, n$, then
$$-2c_i + 2x_i^* - \lambda_i = -2c_i + 2c_i = 0$$

as required. □

Notes:

(1) Many of the results here are the result of joint work with Hannes Leitgeb, which can be found in (Leitgeb & Pettigrew, ms).

(2) For a more negative assessment of the rule, see (Levinstein, 2012).