REFLECTING ON DIACHRONIC DUTCH BOOKS

Michael Rescorla

University of California, Los Angeles

Abstract: *Conditionalization* governs how to reallocate credence in light of new evidence. One prominent argument in favor of Conditionalization holds that an agent who violates it is vulnerable to a *diachronic Dutch book*: a series of acceptable bets offered at multiple times that inflict a sure loss. van Fraassen argues that an agent who violates *the Principle of Reflection* is likewise vulnerable to a diachronic Dutch book. He concludes that agents should conform to both Conditionalization and Reflection. Some authors reply that Reflection is implausible and hence that there must be something wrong with diachronic Dutch book arguments. Other authors try to isolate a principled difference between the Dutch book argument for Conditionalization and the Dutch book argument for Reflection, such that the former argument may succeed even though the latter fails. I pursue a version of this strategy. I contend that, once we properly elucidate the notion of *sure loss*, non-reflectors are not vulnerable to a sure loss. An agent who violates Reflection is not thereby subject to a diachronic Dutch book. Appearances to the contrary result from an unmotivated focus upon an overly narrow set of gambling scenarios.

§1. Conditionalization and Reflection

What are the rational norms that govern changes in credence? The most famous proposed norm is *Conditionalization*, which requires that you respond to evidence E by replacing your old credences P with new credences P_{new} such that:

$$P_{new}(\ .\ ) = P(\ .\ |\ E),$$

where $P(H \mid E)$ is the conditional probability of H given E.[1] Conditionalization figures indispensably in virtually all scientific applications of Bayesian decision theory.

One prominent argument in favor of Conditionalization cites *Dutch books*. A Dutch book is a collection of bets that you regard as acceptable but that inflict a guaranteed net loss upon you. If you are Dutch bookable, then a clever bookie can pump you for money by offering bets that you accept. Thus, Dutch bookability is a very undesirable property. Ramsey (1931) and de Finetti (1937/1980) proved that an agent whose credences violate the probability calculus axioms is Dutch bookable. On that basis, they argued that the probability calculus axioms set norms for rational credal allocation. Lewis extended Dutch book argumentation to Conditionalization. He proved that an agent who follows any update rule other than Conditionalization is Dutch bookable in certain natural learning scenarios. Lewis's proof features a *diachronic* Dutch book (involving bets offered at multiple times). Teller (1973) publicized Lewis's proof, giving credit to Lewis, and Lewis (1999) eventually published his own treatment. Lewis, Teller, and many others deploy Lewis's diachronic Dutch book to argue that one should obey Conditionalization.

Dutch books arguments (DBAs) have grown increasingly controversial over the past few decades (Hájek, 2009; Vineberg, 2011). Diachronic DBAs are especially controversial, partly because they seem to overgeneralize in a problematic way. van Fraassen (1984) articulates a putative requirement on rational credences that he calls the *Principle of Reflection*:

[1] For discussion of how to formulate Conditionalization more carefully, see (Meacham, 2016; Rescorla, 2021).

Reflection: If $P(P_{new}(H) = r) > 0$, then $P(H \mid P_{new}(H) = r) = r$.

Reflection constrains the relation between *second-order credences* (credences about one's own credences) and *first-order credences* (credences about non-credal matters). van Fraassen argues that an agent who violates Reflection (a *non-reflector*) is vulnerable to a diachronic Dutch book. He concludes that credences should conform to Reflection. Yet most philosophers regard Reflection as implausible. There are numerous compelling counterexamples to it (Briggs, 2009), such as cases where I know that my future credences will result from misleading evidence or cases where I know that I will not take my evidence properly into account. Since the diachronic DBA generalizes from Conditionalization to Reflection, and since Reflection looks unacceptable, mustn't there be something amiss with the diachronic DBA for Conditionalization?

Some authors respond by trying to isolate a principled difference between the diachronic DBA for Conditionalization and the diachronic DBA for Reflection, such that the former argument may succeed even though the latter fails (e.g. Briggs, 2009; Mahtani, 2015). I will pursue a version of this strategy. I contend that, once we properly elucidate the notion of *sure loss*, non-reflectors are not vulnerable to a sure loss. An agent who violates Reflection is not thereby subject to a diachronic book that guarantees a loss. Appearances to the contrary result from an unmotivated focus upon an overly narrow set of gambling scenarios. Whether the DBA for Conditionalization succeeds is a further question that I leave unresolved. My thesis is just that Dutch book reasoning does not successfully generalize from Conditionalization to Reflection.

In §§2-3, I critique van Fraassen's DBA for Reflection. I show that the argument goes through only if we restrict attention to scenarios where the bookie has infallible access to the non-reflector's credences. Since we have no reason to impose any such restriction, the argument fails. §§4-5 show that, once we reject dubious infallibility assumptions, a non-reflector who obeys Conditionalization and who satisfies the probability calculus axioms is not vulnerable to a sure loss. In contrast, an agent who follows an update rule other than Conditionalization is vulnerable to a sure loss *even absent any infallibility assumptions*. §6 examines several weakened variants of Reflection. §7 is an appendix that formalizes and proves some key results stated in the main text.

§2. The Dutch book argument for Reflection

Consider an agent with credences P at time t_1 and P_{new} at a later time t_2. Suppose that the agent violates Reflection at t_1. van Fraassen argues that a clever bookie can pump the non-reflector for money as follows. Let

$$P(H \mid P_{new}(H) = r) = s \neq r,$$

and consider the case where $s < r$. At t_1, the bookie offers the non-reflector a conditional bet with net payoffs given by Table 1. In addition, the bookie offers a sidebet on $P_{new}(H)$ with net payoffs given by Table 2. The conditional bet is *acceptable* relative to P, i.e. its expected value relative to P is non-negative. The sidebet is also acceptable relative to P. At t_2, the bookie pursues the following strategy:

(1) If $P_{new}(H) \neq r$, the bookie offers no bet.

If $P_{new}(H) = r$, the bookie offers the bet with payoffs given by Table 3.

The bookie offers the bet from Table 3 precisely when $P_{new}(H) = r$, from which it follows that the bet is acceptable relative to $P_{new}(H)$ whenever the bookie offers it. Table 4 gives net payoffs for the overall gambling scenario. In this table, and throughout the rest of the paper, I assume that the gambling agent has a standing policy of accepting a bet just in case the bet is acceptable. My assumption is legitimate, since the assumed policy is compatible with the dictates of expected utility theory. Net payoffs in Table 4 are everywhere negative. van Fraassen concludes that the bookie inflicts a sure loss upon the non-reflector. The case where $s > r$ is handled similarly, except that all bets are reversed.

INSERT TABLES 1-4 ABOUT HERE

In computing net payoffs, Table 4 makes a crucial assumption: the bookie learns at t_2 whether $P_{new}(H) = r$. Call this *the credal access assumption*. Only under the credal access assumption can the bookie successfully implement strategy (1). Only under the credal access assumption are we entitled to compute net payoffs in accord with Table 4. To see why, suppose that the bookie sometimes forms mistaken beliefs about $P_{new}(H)$. Suppose that the bookie can believe that $P_{new}(H) = r$ even though $P_{new}(H) \neq r$, and suppose that the bookie can believe that $P_{new}(H) \neq r$ even though $P_{new}(H) = r$. Then there are possible outcomes neglected by Table 4. Specifically, there is a possible outcome where $\neg H$ & $P_{new}(H) = r$ but the bookie mistakenly believes that $P_{new}(H) \neq r$. In this outcome, the bookie tries but fails to implement strategy (1). He does not offer the bet from Table 3, so no bet occurs at t_2 to cancel out the non-reflector's winnings from the bets enacted at t_1. Net payoff is

$$s + (r - s)[1 - P(P_{new}(H) = r)],$$

which is positive if $s > 0$ or $P(P_{new}(H) = r) < 1$ and is 0 otherwise. Hence, van Fraassen's proof crucially depends upon the credal access assumption. Absent that assumption, the bookie strategy does not inflict a sure loss upon the non-reflector.

We do not generally have infallible access to one another's credences. I may receive misleading evidence regarding your credences, or I may not take my evidence properly into account. The credal access assumption is false when construed as a general description of gambling interactions among agents. It may be a useful idealizing assumption in certain contexts, such as the game-theoretic study of interactions where deception and error are irrelevant. But we should not confuse a sometimes useful idealizing assumption with a well-motivated global restriction upon gambling transactions.

van Fraassen is commonly taken to have proved a theorem along the following lines: a non-reflector is vulnerable to a diachronic book that inflicts a sure loss. My analysis shows that the theorem is true only if one interprets it using a very restricted notion of "sure loss," on which bets inflict a "sure loss" when they yield a loss in all outcomes where the credal access assumption prevails. Of course, van Fraassen is entitled to stipulate whatever meaning he likes for the phrase "sure loss." But his favored notion of "sure loss" does not seem like a good stipulation once we recognize that the credal access assumption is false. Once we acknowledge that the bookie may mistake the non-reflector's credences at t_2, we see that the bookie strategy described by van Fraassen does not in any natural sense guarantee a loss for the non-reflector. When evaluating whether a bookie strategy inflicts a sure loss, we should consider outcomes where the credal access assumption fails.

van Fraassen might reply that it is only fair to impose the credal access assumption in the present dialectical context, because doing so ensures epistemic parity between the non-reflector and the bookie. Surely the non-reflector knows her own credences at t_2. In particular, the non-reflector t_2 surely knows at t_2 whether $P_{new}(H) = r$. And it is only fair to assume that at t_2 the bookie gains whatever knowledge the non-reflector gains.

The proposed reply assumes that the non-reflector knows at t_2 whether $P_{new}(H) = r$. Call this *the credal introspection assumption*. I reject the credal introspection assumption. As many authors have emphasized (e.g. Roush, 2018; Schwitzgebel, 2008; Williamson, 2000, pp. 93-113, pp. 164-183), introspection is subject to error. We are fallible creatures liable to mistakes in all domains, including introspection. Myriad factors may induce mistaken beliefs about one's own mental states: carelessness, self-deception, exhaustion, inebriation, external manipulation, and so on. To quote Williamson: "Mistakes are always possible. There is no limit to the conclusions into which we can be lured by fallacious reasoning and wishful thinking, charismatic gurus and cheap paperbacks" (2000, p. 94). Credal introspection is particularly error-prone (Christensen, 2007). Do *you* know *your* current credences with complete precision? Can you infallibly determine that your credence in H is a precise number r, as opposed to r – .00000000000001? Surely small mistakes can easily occur. Large mistakes can also occur. For example, a defense attorney who assigns high credence to the proposition that his client is guilty may convince himself that he assigns the proposition low credence. Infallible credal introspection is perhaps a useful idealization in certain contexts, such as contexts where any distorting influence exerted by second-order credence upon first-order credence are negligible. When we debate Reflection, however, our goal is to understand the normative relation between first-order and second-order credences. In this context, it is inadvisable to begin with a blatantly false assumption about how first-order and second-order credences relate. We should allow the possibility of mistaken second-order credences. A compelling argument for Reflection should not invoke the credal introspection assumption.

Reflection fails when the agent recognizes that her future credences may be poorly tuned to reality. van Fraassen's DBA for Reflection flounders once we recognize that second-order credences may be poorly tuned to credal reality. Both Reflection and van Fraassen's DBA for

Reflection enshrine an overly idealized conception of credal formation with little bearing upon the all-too-fallible creatures we know ourselves to be.[2]

§3. Fallible credal introspection

My critique of van Fraassen's DBA hinges upon the possibility of mistaken credal introspection. Many formal epistemologists hold that such mistakes, while possible, are *irrational*. The idea is that a perfectly rational agent will infallibly know her own credences. Influential arguments propounded by Sobel (1987), Uchii (1973), and others allege that faulty credal introspection leaves agents Dutch bookable. So it may seem that fallible second-order credences, rather than circumventing sure loss, actually ensure that undesirable fate.

I now attempt to assuage this worry. I begin by considering two putative norms that are widely discussed, and sometimes implicitly assumed, in formal epistemology:

Accuracy: If $P(P(H) = r) = 1$, then $P(H) = r$.

---

[2] Mahtani (2015) argues that van Fraassen's diachronic book does not inflict a sure loss upon the non-reflector. She defends this position in a different way than I do. She says that a genuine Dutch book should impose a loss *even when we vary the interpretation of the sentences on which agents are betting*. She says that we should hold fixed the interpretation of logical locutions but not the interpretation of any other locutions. To assess whether van Fraassen's diachronic book imposes a sure loss, Mahtani considers alternative possible interpretations for the sentence "$P_{new}(H) = r$", e.g. an interpretation on which "$P_{new}(H) = r$" means that the non-reflector's credence in H is $2r$. She says that van Fraassen's book does not impose a loss under all such interpretations. She concludes that van Fraassen's DBA for Reflection fails. Although I agree with Mahtani's conclusion, I reject her argument for it. When two agents make a bet, they are not betting on an uninterpreted sentence. They are betting on a proposition, or on a sentence *as interpreted a certain fixed way*. Only if one fixes the particular proposition on which one is betting can one say what the bet's net payoff would be given various possible states of the world. Only then can one compute the bet's expected payoff. Only then can one evaluate whether the bet is acceptable and hence whether one should accept it. If sentence S expresses proposition p, then there is certainly a possible scenario where agents use S to express a different proposition q and use S to bet on q rather than p. However, any such scenario is irrelevant to evaluating payoffs from a bet on p itself. A bet on q is different than a bet on p. If we are interested in the ramifications of betting on p, we should hold p fixed. We should hold fixed the interpretation of the sentences that agents use to communicate when they wager with each other. When evaluating whether van Fraassen's book inflicts a sure loss, we should hold fixed the propositions on which the agents are gambling --- proposition H (whatever it may be) and the proposition that $P_{new}(H) = r$. We should consider the results of betting on *those* propositions, not other propositions that might have been expressed by the same sentences. Our question is whether van Fraassen's book inflicts a sure loss *when interpreted as intended*, not whether it inflicts a sure loss *when reinterpreted to involve bets on different propositions*. Thus, Mahtani's analysis does not undermine van Fraassen's DBA.

Confidence: If $P(H) = r$, then $P(P(H) = r) = 1$.

I reject both Accuracy and Confidence. These principles may be useful idealizations for certain purposes, but they are not plausible constraints upon rational credal allocation:

- It is relatively rare to become certain of a false proposition about one's own credences, rarer still for such certainty to be *rational*. Nevertheless, rational violations of Accuracy seem possible when enough misleading evidence accumulates, especially if r is close enough to your true credence in H. For example, you might receive misleading evidence from a brain scanner that normally reveals your credences with tremendous reliability. (Cf. Armstrong, 1963.) Thus, you may in certain circumstances rationally set $P(P(H) = r) = 1$ even though $P(H) \neq r$.
- Since credal introspection is fallible, uncertainty is usually the rationally appropriate stance towards your own credences. Generally speaking, you should maintain second-order credences over possible values for your current first-order credences. For example, you may in some cases rationally set $P(P(H) \in [a, b]) = .9$. When you are uncertain about your own current credences, you violate Confidence.

A rational agent may mistake her own credences or remain uncertain about her own credences.

Uchii (1973) advances Accuracy and Confidence as constraints upon rational second-order credences. He defends his position by offering DBAs for both norms. He contends that someone who violates either norm is vulnerable to a sure loss. His DBA for Accuracy concerns an agent with credences such that

$$P(P(H) = r) = 1 \;\&\; P(H) \neq r.$$

A bookie can offer this agent a bet on $P(H) = r$ with net payoffs given by Table 5. The agent regards the bet as acceptable, so she accepts it. Since $P(H) \neq r$, Uchii contends that the agent has guaranteed herself a net loss. He concludes that agents are Dutch bookable whenever they violate Accuracy. He argues along similar lines for Confidence.

INSERT TABLE 5 ABOUT HERE

If Uchii's DBA were successful, it would pose a serious threat to §2's analysis. It would show that the only way for the non-reflector to evade van Fraassen's diachronic Dutch book is for her second-order credences to expose her to a potential synchronic Dutch book at t_2. However, the DBA is unconvincing. As Roush (2018) emphasizes, the argument holds fixed that $P(H) \neq r$ when assessing possible net payoffs. Whether $P(H) \neq r$ is a contingent matter. It is the very matter about which agents are gambling when they enact the bet from Table 5. To assess prospective net payoffs from this bet, one cannot legitimately hold fixed that $P(H) \neq r$. By comparison, suppose we gamble on whether a fair coin landed heads or tails. The coin landed tails, but you do not know this and have credence .5 that it landed heads. I offer you a bet with net payoff given by Table 6. You regard the bet as acceptable. Does the bet inflict a guaranteed net loss upon you? Of course not. In evaluating whether you are vulnerable to a sure loss, one must consider counterfactual outcomes where the coin landed heads. Even though the coin landed tails, it might have landed heads, so a bet that yields a profit when the coin lands heads does not inflict a sure loss. Similarly, in evaluating whether Uchii's bet imposes a sure loss, one must consider counterfactual outcomes where $P(H) = r$. Even though the agent does not assign credence r to H, she might have done, so a bet that yields net profit when $P(H) = r$ does not inflict a sure loss. Hence, Uchii's DBA for Accuracy is unconvincing. A similar criticism applies to Uchii's DBA for Confidence. For further development of these criticisms, see (Roush, 2018).

INSERT TABLE 6 ABOUT HERE

Another norm that receives attention from formal epistemologists is a synchronic analogue of Reflection:

Self-Respect: If $P(P(H) = r) > 0$, then $P(H \mid P(H) = r) = r$.

Intuitively: you should stand ready to endorse your own current credences. Self-Respect entails Accuracy, as the following argument establishes:

> For any proposition E with $P(E) = 1$, the ratio formula for conditional probabilities entails that $P(H) = P(H \mid E)$. If $P(P(H) = r) = 1$, then $P(H) = P(H \mid P(H) = r)$ and, assuming Self-Respect, $P(H \mid P(H) = r) = r$. Thus, Self-Respect entails that $P(H) = r$ if $P(P(H) = r) = 1$.

van Fraassen (1995), Koons (1992, pp. 20-23), Sobel (1987), and Vickers (2000) claim that credences should conform to Self-Respect.

At first blush, Self-Respect may look more plausible than Reflection. Nevertheless, I agree with Christensen (2007), Lasonen-Aarnio (2015), and Roush (2009) that we should reject it. Counterexamples can arise when you suspect that your credal formation is rationally impaired. To borrow Roush's example, suppose that you witness a crime and on that basis acquire some credence in the proposition J that John is the culprit. Suppose you also have strong evidence that people who are very confident when making eyewitness identifications tend to be overconfident. Then you may rationally set $P(J \mid P(J) = .95) < .95$. As this example illustrates, Self-Respect is not a plausible credal norm. When you suspect that your present credences arose through a defective process, you may rationally question them in a way that violates Self-Respect.

Sobel (1987) defends Self-Respect with a DBA. Consider an agent who violates Self-Respect as follows:

$P(P(H) = r) > 0$.

$P(H \mid P(H) = r) = s < r$.

Sobel's Dutch book involves three bets: a conditional bet on *H*, given by Table 7; a sidebet on $P(H) = r$, given by Table 8; and an unconditional bet on *H*, given by Table 9, which the bookie offers just in case $P(H) = r$. All three bets are acceptable when offered. Net payoff from the overall gambling scenario, given by Table 10, is everywhere negative.

INSERT TABLES 7-10 ABOUT HERE

This is basically just van Fraassen's diachronic book for non-reflectors, adapted to a synchronic setting. The central flaw in the argument, as with van Fraassen's diachronic DBA, is its reliance upon the credal access assumption. Table 10 presupposes that the bookie offers the bet from Table 9 precisely when $P(H) = r$. We should not assume that the bookie can infallibly detect the agent's credence in *H*, because agents do not typically have infallible access to one another's credences. There are possible outcomes where the agent and the bookie both misconstrue the agent's credences. Specifically, consider an outcome where $\neg H$ & $P(H) = r$ but the bookie believes that $P(H) \neq r$. Net payoff in that outcome is

$$s + (r - s)[1 - P(P(H) = r)],$$

which is positive if $s > 0$ or $P(P(H) = r) < 1$ and is 0 otherwise. Thus, Sobel's book does not guarantee a sure loss once we recognize outcomes where the bookie mistakes the agent's credences. For further criticism of the DBA for Self-Respect, see Christensen (2007) and Roush (2018).

Rationality does not require that you infallibly know your own credences. Nor does it bar you from questioning your own credences. Error and self-doubt are rationally permissible. Accuracy, Confidence, and Self-Respect may apply to highly idealized rational beings, but they are not rational requirements upon the credal allocation of ordinary humans.

§4. Dutch book arguments for Conditionalization

I have argued that van Fraassen's DBA fails due to its reliance upon dubious infallibility assumptions. My argument raises some questions:

- Although van Fraassen's diachronic book inflicts a sure loss only given the credal access assumption, is there a *different* diachronic book that inflicts a sure loss absent the credal access assumption? van Fraassen's DBA may not establish that one should conform to Reflection, but does some alternative DBA establish that conclusion?
- To what extent, if at all, does my critique of van Fraassen's DBA for Reflection extend to DBAs for Conditionalization? Do DBAs for Conditionalization also rely upon dubious infallibility assumptions?

The rest of the paper addresses these questions. This section discusses Conditionalization. §5 revisits Reflection.

§4.1 Lewis learning scenarios

I begin by reviewing Lewis's DBA for Conditionalization. Lewis studies an agent who at t_1 has initial credences P over some class of propositions, where P conforms to the probability calculus axioms. Say that $\mathcal{E}$ is a *partition* iff it is a countable set of mutually exclusive, jointly exhaustive propositions such that $0 < P(E) < 1$ for each $E \in \mathcal{E}$. At t_2, the agent learns which partition proposition E is true and as a result adopts new credences P_{new}. More precisely, the agent sets $P_{new}(E) = 1$ for some E, and it is guaranteed that

(2) $P_{new}(E) = 1$ iff E is true.

The agent has determinate dispositions to reallocate credence upon learning that E is true. Let C_E be the agent's credal allocation upon learning that E is true:

(3) If $P_{new}(E) = 1$, then $P_{new} = C_E$.

(2) and (3) entail

(4) If E is true, then $P_{new} = C_E$.

In order for (2) and (3) to hold when E is true, we must also have

(5) $C_E(E) = 1$.

We assume that

(6) C_E conforms to the probability calculus axioms.

Call any learning scenario of this kind a *Lewis learning scenario*.

Suppose that an agent in a Lewis learning scenario violates Conditionalization, setting $C_E(H) \neq P(H \mid E)$ for some proposition H and some partition proposition E. Lewis imagines that the agent faces a bookie with the following two properties:

(7) The bookie learns the true partition proposition at t_2.

(8) The bookie knows at t_1 whether $C_E(H) < P(H \mid E)$ or $C_E(H) > P(H \mid E)$.

Lewis proves that the bookie can pursue a strategy that inflicts a net loss on the agent no matter how the gambling scenario unfolds. Thus, an agent whose update rule differs from Conditionalization is vulnerable to a sure loss in Lewis learning scenarios.

Should we accept the factivity assumptions (2) and (7)? As a general matter, these assumptions may well fail. In any ordinary situation, there is a possibility of *misplaced certainty*, i.e. assignment of credence 1 to a false proposition. I may misread a measuring instrument, or fall victim to deception, or make a computational error, and so on (Rescorla, 2021). Misplaced certainty is possible even when the proposition concerns my own mental states, including my own credences. Most ordinary learning scenarios are not Lewis learning scenarios. (2) and (7) sometimes may be useful idealizations, such as when we wish to ignore possible errors in the evidentiary base for Bayesian inference. As a general matter, they do not support a plausible model of gambling interaction.

Once we lift assumptions (2) and (7), we must acknowledge possible situations where the agent and the bookie both become certain of a false partition proposition. As I show elsewhere (Rescorla, forthcoming), Lewis's diachronic book yields a net profit for the agent in certain such possible situations. Thus, Lewis's DBA for Conditionalization depends in an essential way upon the factivity assumptions (2) and (7).

§4.2 Generalized Lewis learning scenarios

Luckily, we can modify Lewis's proof to avoid any reliance upon factivity assumptions.

Consider an agent whose credences P at t_1 satisfy the probability calculus axioms and who becomes certain at t_2 of a proposition E drawn from partition $\mathcal{E}$. The agent has determinate dispositions to reallocate credence in light of her newfound certainty in E, yielding a new credal allocation C_E. We still assume (3), (5), and (6). However, we no longer assume (2) and (4). We do not assume that the agent learns the true partition proposition. Intuitively: the agent receives some partition proposition E as evidence, but E may not be true. Call any learning scenario of this kind a *generalized Lewis learning scenario*. In a generalized Lewis learning scenario, the agent may become certain of a partition proposition E even though a different partition proposition is true.

Having broadened attention to generalized Lewis learning scenarios, we must reconsider our treatment of gambling. We still assume that the bookie has a strategy for deciding what bet to offer at t_2, depending on the partition proposition of which he becomes certain. However, it no longer seems appropriate to assume (7). If a bookie learns the true partition proposition but the agent has no such luck, then it hardly demonstrates irrationality on the agent's part if the bookie can inflict a sure loss. It now seems appropriate to assume merely that agent and bookie both become certain of the same partition proposition E at t_2, where E may be false. Under this revised conception, a gambling interaction proceeds much as under Lewis's conception. On both conceptions, there is a partition proposition that the agent and the bookie come to invest with complete certainty. The key difference is that under our revised conception the agent and bookie may be mistaken: the partition proposition may be false.

Many possible outcomes ignored by Lewis now require consideration. See Figure 1. When assessing what counts as a diachronic Dutch book, Lewis confines attention to outcomes where both agents learn the true partition proposition. In terms of Figure 1: he confines attention to the diagonal cells. Having broadened attention to generalized Lewis learning scenarios, we should consider all cells, not merely diagonal cells. Whereas Lewis demands that a diachronic Dutch book inflict a negative net payoff along the diagonal, I instead demand that a diachronic Dutch book inflict a negative payoff on all cells. Call a diachronic Dutch book understood Lewis's way a *factive diachronic Dutch book.* Call a diachronic Dutch book understood my way a *non-factive diachronic Dutch book.*

INSERT FIGURE 1 ABOUT HERE

Suppose that an agent in a generalized Lewis learning scenario violates Conditionalization, setting $C_E(H) \neq P(H \mid E)$ for some proposition H and some partition proposition E. As I show in (Rescorla, forthcoming), one can mount a non-factive diachronic Dutch book against this agent. Without loss of generality, consider the case where $C_E(H) < P(H \mid E)$. Assuming (8), the bookie may proceed as follows. At t_1, he offers the conditional bet given by Table 11 and the sidebet on E given by Table 12. At t_2, both agents become certain of some partition proposition E^*. No matter the partition proposition of which they become certain, the bookie offers the conditional bet given by Table 13. All three bets are acceptable when offered. Net payoff, given by Table 14, is everywhere negative. Thus, an agent who employs any update rule other than Conditionalization is vulnerable to a sure loss in generalized Lewis learning scenarios.

INSERT TABLES 11-14 ABOUT HERE

Skyrms (1987) proves a *converse Dutch book theorem*: a conditionalizer whose credences conform to the probability calculus axioms is invulnerable to a sure loss in Lewis learning scenarios. If you start with credences that satisfy the probability calculus axioms, and if you update your credences by conditionalizing on the revealed partition proposition, then no diachronic Dutch book can be rigged against you. Skyrms's theorem easily extends to a converse Dutch book theorem for generalized Lewis learning scenarios: a conditionalizer who obeys the probability calculus axioms is immune to a sure loss in generalized Lewis learning scenarios. One cannot rig a non-factive Dutch book against a conditionalizer who obeys the probability calculus axioms. For a formal statement of the converse non-factive Dutch book theorem, see the appendix (§7.1). The converse non-factive theorem and the non-factive Dutch book from the previous paragraph jointly establish Conditionalization as the unique credal reallocation strategy that avoids non-factive Dutch books in generalized Lewis learning scenarios.

§5. Failing to reflect

To assess how non-factive Dutch books bear upon Reflection, we must incorporate higher-order credences into a formal model of credal allocation. There is a substantial economics literature, initiated by the seminal work of Harsanyi (1967), that pursues formal modeling of higher-order credence. The topic has also received some discussion among philosophers (e.g. Campbell-Moore, 2015; Gaifman, 1998; Skyrms, 1980), although not as much discussion as one might expect given how much attention philosophers have lavished upon Reflection. In the appendix (§7.2), I give a relatively simple formal framework that can model a wide range of higher-order credences, including the agent's current credence in the proposition $P_{new}(H) = r$.

Crucially, the converse non-factive Dutch book theorem persists *even after we incorporate higher-order credences into the formal model*. The theorem has dramatic consequences regarding Reflection. Consider a non-reflector who obeys Conditionalization and who satisfies the probability calculus axioms. Is she vulnerable to a sure loss in generalized Lewis learning scenarios? The converse non-factive Dutch book theorem entails that she is not. One cannot rig a non-factive Dutch book against her, even if her initial credences violate

Reflection. Hence, a conditionalizer who satisfies the probability calculus axioms is immune to a sure loss in generalized Lewis learning scenarios *whether or not she satisfies Reflection*.

We may instructively compare the fate of non-reflectors in Lewis learning scenarios and generalized Lewis learning scenarios. If you violate Reflection, then there exist Lewis learning scenarios in which you simply cannot conditionalize. More specifically, suppose you set

$$P(H \mid P_{new}(H) = r) = s \neq r,$$

and suppose you participate in a Lewis learning scenario involving a partition that contains the proposition $P_{new}(H) = r$. Whatever update rule you employ, the update rule cannot be Conditionalization. To see why, suppose that partition proposition $P_{new}(H) = r$ is revealed at t_2, leading you to form new credences P_{new}. Then you cannot arrive at your new credence $P_{new}(H)$ by conditionalizing on the revealed partition proposition $P_{new}(H) = r$. If you were to conditionalize on $P_{new}(H) = r$, you would set

$$P_{new}(H) = P(H \mid P_{new}(H) = r) = s \neq r,$$

which conflicts with the assumption that $P_{new}(H) = r$. Your credences P_{new} must instead result from an update rule C such that

$$C_{P_{new}(H)=r}(H) = r \neq P(H \mid P_{new}(H) = r)\,.$$

Any such update rule conflicts with Conditionalization. However exactly you arrive at your new credences P_{new} in response to learning that $P_{new}(H) = r$, you cannot do so by conditionalizing upon partition proposition $P_{new}(H) = r$.

The situation is very different once we consider broaden attention from Lewis learning scenarios to generalized Lewis learning scenarios. The non-reflector *can* conditionalize in all generalized Lewis learning scenarios. Specifically, she can conditionalize on the partition proposition $P_{new}(H) = r$, setting

$P_{new}(H) = P(H \mid P_{new}(H) = r\ ) = s \neq r.$

If she does so, she will have misplaced certainty about her own credences. She will set

$P_{new}(P_{new}(H) = r\ )) = 1$

even though

$P_{new}(H) = s \neq r.$

See Figure 2. A conditionalizing non-reflector is pushed off the grey cells of Figure 2. van Fraassen's proof works by forcing the non-reflector onto the grey cells, which bars her from using Conditionalization as her update rule and thereby exposes her to sure loss.[3]

INSERT FIGURE 2 ABOUT HERE

Let us examine in more detail how a conditionalizing non-reflector fares against van Fraassen's bookie in generalized Lewis learning scenarios. As always, I focus without loss of generality on the case where $P(H \mid P_{new}(H) = r) = s < r$. Suppose the non-reflector participates in a generalized Lewis learning scenario that includes $P_{new}(H) = r$ as a partition proposition. At t_1, the bookie offers the bets from Tables 1 and 2. At t_2, the bookie tries to pursue strategy (1). He offers the bet from Table 3 in precisely those situations where he and the non-reflector become certain of $P_{new}(H) = r$ at t_2. So he inadvertently conforms to the following rule:

(9) If $P_{new}(P_{new}(H) = r) \neq 1$, the bookie offers no bet.

If $P_{new}(P_{new}(H) = r) = 1$, the bookie offers the bet with payoffs given by Table 3.

[3] van Fraassen (1995) argues that someone who satisfies Conditionalization must also satisfy Reflection. As Weisberg (2007) observes, the argument only goes through if we assume that the agent is certain she will conditionalize at t_2 and is certain she will satisfy Confidence at t_2. There is no reason why we should grant van Fraassen these assumptions. A rational agent need not be certain that, at some future time, she will conditionalize or will satisfy Confidence. The formal framework presented in the appendix shows that a non-reflector who satisfies the probability calculus axioms can obey Conditionalization.

When $P_{new}(P_{new}(H) = r) = 1$, the non-reflector conditionalizes on $P_{new}(H) = r$ and sets

$$P_{new}(H) = P(H \mid P_{new}(H) = r) = s < r.$$

An easy computation shows that the bet from Table 3 is unacceptable at t_2 in light of the new credence assigned to H. The non-reflector therefore rejects the bet as unacceptable. As a result the bet is not enacted at t_2 *whether or not the non-reflector becomes certain of* $P_{new}(H) = r$ *at* t_2. Net payoff is determined entirely by Tables 1 and 2. In particular, net payoff is

$$s + (r - s)[1 - P(P_{new}(H) = r)] > 0.$$

whenever $\neg H$ & $P_{new}(H) = r$. Thus, van Fraassen's book does not inflict a sure loss upon the conditionalizing non-reflector.

To assess van Fraassen's book in the context of generalized Lewis learning scenarios, I have analyzed a rather unusual learning situation: namely, a situation in which the agent conditionalizes on the proposition $P_{new}(H) = r$. Becoming certain that your new credence in H is r is a strange way of gaining new credences. Agents do not typically conditionalize on second-order propositions about their own credences. My analysis shows that, *even if we allow such unusual learning scenarios*, van Fraassen's book does not inflict a sure loss.

The converse non-factive Dutch book theorem also applies to more typical learning scenarios in which agents conditionalize on first-order rather than second-order propositions. The theorem establishes a general conclusion that extends beyond van Fraassen's specific book and beyond unusual situations in which agents conditionalize on second-order propositions about their own credences. From the theorem, we gain blanket assurance that a conditionalizer who conforms to the probability calculus axioms is immune to a sure loss in *any* generalized Lewis learning scenario. This immunity persists even when the agent's initial credences are deviant or ill-advised. It certainly persists when her initial credences violate Reflection. Hence, one should not hope to replace van Fraassen's book with an alternative book that imposes a sure loss upon non-reflectors. The converse non-factive Dutch book theorem tells us that no such book exists.

Of course, there are many possible learning situations that are not generalized Lewis learning scenarios: situations where credal mass redistributes over the partition, with no single proposition receiving all probability mass (Jeffrey, 1983); situations involving memory loss or the threat of memory loss (Arntzenius, 2003); situations where the agent acquires new concepts and thereby comes to entertain previously unavailable propositions (Lewis, 1999); and so on. How do non-reflectors fare in such situations? This is an interesting question. We need not answer it in order to pinpoint the mistake underlying van Fraassen's DBA for Reflection. By showing that conditionalizing non-reflectors are immune to a sure loss in generalized Lewis learning scenarios, I have shown that violations of Reflection *per se* do not induce vulnerability to a sure loss.

A crucial asymmetry emerges between Conditionalization and Reflection. Agents who employ any update rule other than Conditionalization are vulnerable to a sure loss in generalized Lewis learning scenarios, while conditionalizers are immune to that unpleasant prospect. In particular, conditionalizing non-reflectors are immune to a sure loss in generalized Lewis learning scenarios. Non-reflectors are not (by virtue of being non-reflectors) vulnerable to a sure loss in generalized Lewis learning scenarios. Once we consider a suitably broad range of learning scenarios, non-reflectors do not look remotely Dutch bookable. I conclude that there can be no compelling DBA for Reflection. Whether the DBA for Conditionalization succeeds is a trickier question that I am not addressing. My thesis is just that the DBA for Conditionalization does not generalize into a compelling argument for Reflection.

§6. Weakened reflection principles

There are numerous ways one might weaken Reflection by altering its antecedent. My analysis from §5 extends to many of these weakened reflection principles, although not all.[4]

Consider the following weakened reflection principle, which Mahtani (2012) articulates and then critiques:

> ***Adapted Reflection***: If $P(P_{new}(H) = r) > 0$, and $P_{new}(P_{new}(H) = r) = 1$ if $P_{new}(H) = r$, and $P_{new}(P_{new}(H) \neq r) = 1$ if $P_{new}(H) \neq r$, then $P(H \mid P_{new}(H) = r) = r$.

As Mahtani emphasizes, Adapted Reflection looks no more plausible than Reflection. Counterexamples to Reflection are readily converted into counterexamples to Adapted Reflection, by stipulating that the non-reflector becomes certain at t_2 as to whether $P_{new}(H) = r$ and that her certainty is not misplaced. A similar but slightly different principle is:

> ***Faithful Reflection***: If $P(P_{new}(H) = r) > 0$, and $P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r) = 1$, then $P(H \mid P_{new}(H) = r) = r$.

Faithful Reflection also seems quite implausible.

Focusing on Faithful Reflection, consider an agent such that $P(P_{new}(H) = r) > 0$, $P(H \mid P_{new}(H) = r) = s < r$, and

(10) $P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r) = 1$.

Let us assume that the bookie is certain of $P_{new}(H) = r$ iff the non-reflector is certain of it. Let us also assume that the bookie offers the bets from Tables 1 and 2 at time t_1 and implements (9) at t_2. Under these assumptions, Table 15 gives net payoffs for the overall gambling scenario. The

---

[4] In the extremal case where $P(P_{new}(H) = r) = 1$, Reflection entails a weaker principle that I will call ***Extremal Reflection***: If $P(P_{new}(H) = r) = 1$, then $P(H) = r$. When we assess Extremal Reflection, we can no longer consider a generalized Lewis learning scenario that includes $P_{new}(H) = r$ as a partition proposition, because partition propositions must have prior credence < 1. So §5's argument does not apply to Extremal Reflection. Nevertheless, one can show through kindred reasoning that a conditionalizer who violates Extremal Reflection is not thereby vulnerable to a sure loss. For reasons of space, I must leave this claim undefended.

grey cells of Table 15 are those where (10) prevails. The final outcome falls in one of the grey cells, where payoff is negative. Apparently, then, we can inflict a sure loss on someone who violates Faithful Reflection. Does a compelling DBA for Reflection ensue?

INSERT TABLE 15 ABOUT HERE

To see where this reasoning goes astray, let us reflect more carefully upon what counts as a "sure loss." Suppose I have credence .2 *that Hesperus is Phosphorus*. I therefore accept a bet with net payoff –.8 if Hesperus is Phosphorus and .2 if Hesperus is not Phosphorus. As Kripke (1980) has shown, it is metaphysically necessary that Hesperus is Phosphorus. Thus, I lose money in all metaphysically possible outcomes. Does it follow that I am vulnerable to a sure loss? No. From *my* viewpoint, I am not guaranteed to lose. From *my* viewpoint, there are epistemically possible outcomes where I win.[5] When evaluating whether a book guarantees a sure loss, we should consider which outcomes are possible *from the agent's perspective*. As Easwaran (2011, p. 315) puts it, "the notion of 'guarantee' here must be subjective." The question is not whether a loss is guaranteed by metaphysical necessity but whether *the agent* regards a loss as guaranteed.

We can apply these lessons to Table 15. When we evaluate whether a book inflicts a sure loss, we must consider all outcomes that are epistemically possible for the agent at time t_1. In the present case, there are epistemically possible outcomes where payoff is nonnegative. Credal introspection is fallible. Even if we assume that (10) is true, there are possible outcomes where it

[5] What is "epistemic possibility"? The literature offers various candidate notions (e.g. Huemer, 2007; Chalmers, 2012, p. 470). Given the Hesperus-Phosphorus example, any complete treatment of DBAs must grapple with these issues. For present purposes, I leave epistemic possibility unelucidated. I think that any reasonable notion of epistemic possibility will support my conclusions in the main text.

fails. Specifically, there is a possible outcome where $\neg H$ & $P_{new}(H) = r$ but $P_{new}(P_{new}(H) = r) \neq 1$. Assuming that the non-reflector is remotely like us, she recognizes this outcome as possible. *From her viewpoint*, it is a live possibility that she lands in the cell where $\neg H$ & $P_{new}(H) = r$ but $P_{new}(P_{new}(H) = r) \neq 1$. Net payoff is nonnegative in that cell. So the diachronic Dutch book does not inflict a sure loss upon her *in the appropriate sense of "sure loss."* Even if a loss is guaranteed by (10), the non-reflector does not regard it as guaranteed. From her perspective, there are possible outcomes where she does not lose. Thus, van Fraassen's diachronic Dutch book does not inflict a sure loss *in any appropriate sense of "sure loss"* upon someone who violates Faithful Reflection. A similar diagnosis applies to Adapted Reflection.[6]

In response, one might weaken Faithful Reflection by building in an assumption that cells with nonnegative payoff are not epistemically possible:

> ***Confident Reflection***: If $P(P_{new}(H) = r) > 0$, and $P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r) = 1$, and it is epistemically necessary that $P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r) = 1$, then $P(H \mid P_{new}(H) = r) = r$.

Here I use

> It is epistemically necessary that *p*

as shorthand for

---

[6] Briggs (2009) proposes a *suppositional test* for differentiating between convincing and unconvincing DBAs. The basic idea is that bets reveal irrationality when they inflict a net loss not just in possible worlds where the agent accepts the bets but also in certain other possible worlds. Briggs claims that, according to the suppositional test, the Dutch book from Lewis's DBA reveals irrationality but the Dutch book from van Fraassen's DBA does not reveal irrationality. Mahtani (2012) argues convincingly that Briggs's suppositional test, even if it yields plausible verdicts for Conditionalization and Reflection, yields an implausible verdict for Adapted Reflection. As Mahtani emphasizes, the suppositional test confines attention to possible worlds where the partition proposition of which the agent becomes certain at t_2 is true. If the agent violates Adapted Reflection, then the relevant possible worlds fall within the grey cells of Table 15, where net payoff is negative. So the suppositional test mistakenly entails that agents who violate Adapted Reflection are irrational. Mahtani concludes, and I agree, that Briggs's suppositional test does not successfully differentiate between convincing and unconvincing DBAs. To explain why the DBA for Adapted Reflection fails, one must consider the white cells of Table 15, cells which the suppositional test dismisses as irrelevant.

It is not epistemically possible that $\neg p$,

and I implicitly relativize these modal operators to an agent at a moment of time. The precise content of Confident Reflection depends upon how one elucidates epistemic possibility. Lacking further elucidation, it is not evident how Confident Reflection interfaces with formal modeling of credence, including the converse non-factive Dutch book theorem presented in the appendix. Still, one can mount a strong *prima facie* case that agents who violate Confident Reflection are Dutch bookable. For any such agent, the white cells of Table 15 are not epistemically possible. (10) is true in the grey cells of Table 15, false in the white cells. van Fraassen's diachronic book inflicts a net loss in grey cells. So van Fraassen's diachronic book apparently inflicts a loss upon the agent in *all* epistemically possible outcomes.

Let us grant, for the sake of argument, that one can mount a compelling DBA for Confident Reflection. Would that be worrisome? Speaking for myself, I find Confident Reflection fairly anodyne. We are fallible even when it comes our own credences, so any rational agent should recognize that (10)'s right-to-left direction may fail. She should also recognize at t_1 that she may remain uncertain at t_2 about her own credences, in which case (10)'s left-to-right direction fails. Thus, (10) is not epistemically necessary for any agent remotely like us. When violations of (10) are epistemically possible for an agent, the agent trivially satisfies Confident Reflection. Hence, Confident Reflection is an extremely weak norm. It applies in a non-trivial way only to agents who have perfect introspective access to their own credences *and* who brook no possibility that this perfect introspective access will fail. Such agents do not exist. If they did exist, they would differ profoundly from us. I therefore doubt that Confident Reflection is subject to damning counterexamples of the sort that plague full-blown Reflection.[7]

---

[7] In the same ballpark as Confident Reflection lies a norm that I will call ***Complacent Reflection***: If $P(P_{new}(H) = r) > 0$, and $P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r) = 1$, and $P(P_{new}(H) = r$ iff $P_{new}(P_{new}(H) = r)) = 1$, then $P(H \mid P_{new}(H) = r) = r$.

There are many other weakened reflection principles one might consider. I question whether any substantial, well-motivated constraint on the relation between first-order and second-order credences in humans or human-like creatures would result.[8] Reflection itself is a very implausible norm and is widely recognized as such. The reason it has received so much attention over the past few decades is that it seems to derive support from an argumentative strategy widely deployed to support Conditionalization. I have shown that the argumentative strategy, when properly developed, does not actually extend from Conditionalization to Reflection. The strategy extends in a convincing way only if we weaken Reflection so much that it loses all interesting applicability, as with Confident Reflection. I submit that Reflection and its weakened variants are red herrings, serving mainly to divert attention from more fruitful lines of inquiry into rational credal allocation.

§7. Appendix: formal modeling of Dutch books and higher-order credence

This appendix formalizes and proves key claims from the paper. I am concerned throughout with a *probability space* $(\Omega, \mathfrak{F}, P)$, where Ω is a set, $\mathfrak{F}$ is a σ-field over Ω, and P is a probability measure on $\mathfrak{F}$. Elements of Ω are *outcomes*. Elements of $\mathfrak{F}$ are *events*. Heuristically,

---

Since violations of (10) may be epistemically possible for an agent who violates Complacent Reflection, my analysis of Faithful Reflection carries over to Complacent Reflection. But suppose for the sake of argument that one could mount a compelling DBA for Complacent Reflection. Would that be so terrible? As Weisberg (2007) notes, it is hardly a norm of rationality that one should be certain of (10). On the contrary, certainty in one's future introspective certainty and infallibility seems foolhardy in most circumstances. Even if we were to accept Complacent Reflection, it would not follow that an agent should set $P(H \mid P_{new}(H) = r) = r$. The agent could just as well conform to Complacent Reflection by abandoning her certainty in (10).

[8] Pettigrew (forthcoming) gives a non-factive Dutch book argument for two principles that he calls the *Weak General Reflection Principle* and the *Strong General Reflection Principle*. Weak General Reflection and Strong General Reflection both constrain the relation between your present credences and your possible future credences. However, neither principle constrains the relation between your *present first-order credences* and your *present second-order credences*. For that reason, neither principle entails van Fraassen's original Reflection principle or any of the weaker variants I have discussed. Thus, Pettigrew's position is compatible with mine.

one can interpret outcomes as possible worlds and events as sets of possible worlds. I use $(\Omega, \mathscr{F}, P)$ to model an agent's credences at t_1.

§7.1 Dutch books formalized

We want to model how the agent updates her credences at t_2 in light of new certainties. A *partition* is a countable set $\mathscr{E}$ such that

$$\bigcup_{E \in \mathscr{E}} E = \Omega$$

$$E \in \mathscr{F} \qquad \textit{for all } E \in \mathscr{E}$$

$$E \cap E^* = \varnothing \qquad \textit{for all } E, E^* \in \mathscr{E}$$

$$0 < P(E) < 1 \qquad \textit{for all } E \in \mathscr{E}.$$

Let $\Delta(\mathscr{F})$ be the set of probability measures over the measurable space $(\Omega, \mathscr{F})$. An *update rule for* $(\Omega, \mathscr{F}, \mathscr{E})$ is a function $C: \mathscr{E} \rightarrow \Delta(\mathscr{F})$ satisfying the condition:

(11) $C(E)(E) = 1.$

I notate $C(E)$ as C_E, so that (11) becomes:

(12) $C_E(E) = 1.$

(12) reflects our intuitive starting point: C_E gives the agent's credences at t_2 in light of her newfound certainty in E.

The agent faces a bookie who may offer bets at t_1 and t_2. Multiple bets offered at a given time can be combined into a single bet, so we may legitimately assume that the bookie offers a single bet at t_1 and a single bet at t_2. Following standard practice in probability theory, I formalize a bet as a random variable. More specifically, a bet is an $\mathscr{F}$-measurable function $X: \Omega \rightarrow \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ is the extended real line. $X(\omega)$ is the net payoff for outcome ω. The random variable whose value is everywhere 0 models a situation where the bookie offers no bet. A bet X is *acceptable relative to P* iff its expected value with respect to P is non-negative:

$$\int XdP \geq 0.$$

Acceptability relative to P formalizes the agent's acceptability verdicts at t_1. At t_2, the agent acquires credences C_E based on her newfound certainty in E. Say that X is *acceptable given E* iff its expected value with respect to C_E is non-negative:

$$\int XdC_E \geq 0.$$

Acceptability given E formalizes the agent's acceptability verdicts at t_2.

The agent and the bookie become certain of the same partition proposition at t_2. The bookie has a strategy for deciding which bet to offer at t_2 depending on this proposition. We model the bookie's strategy using a function Y: $\mathfrak{E} \times \Omega \rightarrow \overline{\mathbb{R}}$ such that, for all E, $Y(E, .)$: $\Omega \rightarrow \overline{\mathbb{R}}$ is a random variable. Call any such function a *bookie strategy*. Notate $Y(E, .)$ as Y_E.

With these definitions in place, we may formalize Dutch bookability. I introduce two formal notions, corresponding respectively to the intuitive notions *sure loss in Lewis learning scenarios* and *sure loss in generalized Lewis learning scenarios*. A *factive Dutch book for* (Ω, $\mathfrak{F}$, P) *and update rule C* is an ordered pair (X, Y) such that

(a) X is a bet that is acceptable relative to P.

(b) Y is a bookie strategy.

(c) Y_E is acceptable given E.

(d) For all $E \in \mathfrak{E}$ and all $\omega \in E$, $X(\omega) + Y_E(\omega) < 0$.

A factive Dutch ensures a net loss in all situations where the agent and the bookie become certain of the true partition proposition. A *non-factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and update rule* C is an ordered pair (X, Y) that satisfies (a)-(c) along with

(d*) For all $E \in \mathfrak{E}$ and all $\omega \in \Omega$, $X(\omega) + Y_E(\omega) < 0$.

A non-factive Dutch book ensures a net loss in all situations, even situations where the agent and bookie acquire misplaced certainty in a partition proposition.

Factive Dutch Book Theorem for Conditionalization (Lewis): *Let* $(\Omega, \mathfrak{F}, P)$ *be a probability space, let* $\mathfrak{E}$ *be a partition, and let* C *be an update rule for* $(\Omega, \mathfrak{F}, \mathfrak{E})$*. If there exists* $E \in \mathfrak{E}$ *such that* $C_E \neq P(\, . \mid E)$*, then there exists a factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and* C*.*

Converse Factive Dutch Book Theorem for Conditionalization (Skyrms): *Let* $(\Omega, \mathfrak{F}, P)$ *be a probability space, let* $\mathfrak{E}$ *be a partition, and let* C *be an update rule for* $(\Omega, \mathfrak{F}, \mathfrak{E})$*. If* $C_E = P(\, . \mid E)$ *for all* $E \in \mathfrak{E}$*, then there does not exist a factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and* C*.*

Non-factive Dutch Book Theorem for Conditionalization: *Let* $(\Omega, \mathfrak{F}, P)$ *be a probability space, let* $\mathfrak{E}$ *be a partition, and let* C *be an update rule for* $(\Omega, \mathfrak{F}, \mathfrak{E})$*. If there exists* $E \in \mathfrak{E}$ *such that* $C_E \neq P(\, . \mid E)$*, then there exists a non-factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and* C*.*

Converse Non-factive Dutch Book Theorem for Conditionalization: *Let* $(\Omega, \mathfrak{F}, P)$ *be a probability space, let* $\mathfrak{E}$ *be a partition, and let* C *be an update rule for* $(\Omega, \mathfrak{F}, \mathfrak{E})$*. If* $C_E = P(\, . \mid E)$ *for all* $E \in \mathfrak{E}$*, then there does not exist a non-factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and* C*.*

The proof of the third theorem formalizes the Dutch book given by Tables 11-14. A non-factive Dutch book is a factive Dutch book, so the third theorem entails the first and the second theorem entails the fourth. For proof of the second theorem, see (Skyrms, 1987; Rescorla 2018). See (Rescorla, forthcoming) for extensive discussion of all four theorems.[9]

§7.2 Second-order credences formalized

To apply §7.1's theorems to Reflection, we must incorporate higher-order credences into our formal model. As Skyrms (1987) suggests, the most natural strategy consistent with modern probability theory is to deploy $\Delta(\mathfrak{F})$: the set of probability measures on $(\Omega, \mathfrak{F})$. We interpret members of $\Delta(\mathfrak{F})$ as possible credal allocations over $\mathfrak{F}$ at t_2. To define second-order credences using $\Delta(\mathfrak{F})$, we must impose measurable structure upon $\Delta(\mathfrak{F})$. Luckily, probability theory provides us with the needed tools. For any $H \in \mathfrak{F}$, define $\Pi_H : \Delta(\mathfrak{F}) \rightarrow [0, 1]$ by

$$\Pi_H(\mu) = \mu(H) .$$

Let $\mathfrak{I}$ be the smallest σ-field over $\Delta(\mathfrak{F})$ such that, for each $H \in \mathfrak{F}$, the function Π_H is $\mathfrak{I}$-measurable. $(\Delta(\mathfrak{F}), \mathfrak{I})$ is the *measurable space of probability measures on* $(\Omega, \mathfrak{F})$. See (Fristedt & Gray, 1997, p. 413) for discussion of $(\Delta(\mathfrak{F}), \mathfrak{I})$.

---

[9] The converse factive theorem proved by Skyrms (1987) is somewhat weaker than the converse factive theorem stated in the main text. Skyrms confines attention to a subclass of bookie strategies: namely, strategies that at t_2 offer bets on finitely many events drawn from $\mathfrak{F}$. (Rescorla, 2018) proves a generalized theorem that extends Skyrms's result to arbitrary bookie strategies Y. The generalized theorem extends Skyrms's theorem in another respect as well: Skyrms's theorem only covers cases where $P(E) > 0$, while the generalized theorem covers numerous cases where $P(E) = 0$. The extra coverage is enabled by Kolmogorov's (1933/1956) theory of conditional probability, which goes beyond the ratio formula to analyze conditional probability even when $P(E) = 0$. In the present paper, I confine attention to cases where $P(E) > 0$, as codified by the assumption that partition propositions have non-zero probability.

We want to model the agent's credences, including first-order credences over events in $\mathfrak{F}$, second-order credences over first-order credences at t_2, and mixed-order credences. $(\Delta(\mathfrak{F}), \mathfrak{I})$ taken on its own suffices to model second-order credences. For example, the event

$$\{\mu: \mu(H) = r\}$$

can serve as a formal proxy for the proposition that $P_{new}(H) = r$. More generally, the event

$$\{\mu: \mu(H) \in [a, b]\} \qquad a, b \in \mathbb{R}$$

corresponds to the proposition that $a \leq P_{new}(H) \leq b$.

To model mixed-order credences, we must find some way of bringing events drawn from $\mathfrak{I}$ into contact with events drawn from $\mathfrak{F}$. An elegant solution, introduced by Harsanyi's (1967) work on games of incomplete information, is to posit a measurable mapping $T: \Omega \rightarrow \Delta(\mathfrak{F})$. This posit captures the intuitive idea that the agent's credence at t_2 is itself a random variable that depends upon the state of the world. $\mathfrak{I}$ is generated by the events

$$\Pi_H^{-1}(-\infty, a] \qquad H \in \mathfrak{F},\ a \in \mathbb{R},$$

so that T is measurable iff

$$\mathrm{T}^{-1}\Pi_H^{-1}(-\infty, a] \in \mathfrak{F} \qquad \textit{for all } H \in \mathfrak{F},\ a \in \mathbb{R}.$$

Moreover,

$$\Pi_H^{-1}(-\infty, a] = \{\mu: \mu(H) \leq a\},$$

so that

$$T^{-1}\Pi_H^{-1}(-\infty, a] = \{\omega: T(\omega)(H) \leq a\}.$$

With these points in mind, we take our basic object of study to be a quadruple $(\Omega, \mathfrak{F}, P, T)$, where $(\Omega, \mathfrak{F}, P)$ is a probability space and $T: \Omega \rightarrow \Delta(\mathfrak{F})$ is such that

$$\{\omega: T(\omega)(H) \leq a\} \in \mathfrak{F} \qquad \textit{for all } H \in \mathfrak{F},\ a \in \mathbb{R}.$$

Call any such $(\Omega, \mathfrak{F}, P, T)$ a *higher-order probability space*. See Gaifman (1988) and Samet (1999) for discussion of higher-order probability spaces.[10]

Given the use to which I am putting $(\Omega, \mathfrak{F}, P, T)$, it makes sense to introduce some clarifying notation. For any $H \in \mathfrak{F}$, let the expression

$$P_{new}(H) = r$$

serve as a name for the event

(13) $\{\omega \colon T(\omega)(H) = r\}$.

This notation registers that (13) is our formal proxy for the proposition that the agent's credence in H is r at t_2. By assumption, (13) belongs to $\mathfrak{F}$. Thus, P is well-defined on (13). In other words, there is a unique x such that

$$P(P_{new}(H) = r) = x.$$

We may also consider events such as

(14) $G \cap P_{new}(H) = r$ $\qquad G, H \in \mathfrak{F}$.

Since (14) belongs to $\mathfrak{F}$, P is well-defined on (14). As a special case of (14), consider

$$H \cap P_{new}(H) = r \qquad H \in \mathfrak{F}.$$

Using the ratio formula, the conditional probability $P(H \mid P_{new}(H) = r)$ is well-defined so long as $P(P_{new}(H) = r) > 0$.

The notion of a higher-probability space does not require $P(H \mid P_{new}(H) = r) = r$. Here is a simple example. Let $S = \{\omega_1, \omega_2\}$ and let $\mathfrak{S}$ be the powerset of S. We specify a probability measure over $\mathfrak{S}$ by specifying the probabilities assigned to $\{\omega_1\}$ and $\{\omega_2\}$. Let P be the measure

---

[10] Gaifman and Samet both assume formalized versions of Reflection. As I have urged in the main text, Reflection is not a rational constraint upon second-order credences. Thus, one should not assume Reflection when developing the notion of a higher-order probability space in full generality.

that assigns value ¼ to $\{\omega_1\}$ and ¾ to $\{\omega_2\}$. Notate this measure as $\langle$¼, ¾$\rangle$. Using the same notation, define T by:

$$T(\omega_1) = \langle 0, 1\rangle$$

$$T(\omega_2) = \langle 1, 0\rangle.$$

Then $(S, \mathbb{S}, P, T)$ is a higher-order probability space. Intuitively, $(S, \mathbb{S}, P, T)$ models a situation where at t_2 the agent becomes certain of the wrong event. Note that

$$P(\{\omega_1\} \mid P_{new}(\{\omega_1\}) = 1) = P(\{\omega_1\} \mid \{\omega\text{: } T(\omega)(\{\omega_1\}) = 1\}) = P(\{\omega_1\} \mid \{\omega_2\}) = 0.$$

So $(S, \mathbb{S}, P, T)$ violates Reflection. One can build far more elaborate higher-order probability spaces that illustrate the same point.

Given a higher-order probability space $(\Omega, \mathfrak{F}, P, T)$, we can iterate the construction of higher-order probabilities. Consider the third-order event:

$$\{\omega\text{: } T(\omega)(P_{new}(H) = r) = 1\},$$

which given our notational conventions may be notated as

$$P_{new}((P_{new}(H) = r)) = 1.$$

Another notable event is

$$(P_{new}(H) = r \cap P_{new}((P_{new}(H) = r)) = 1) \cup (P_{new}(H) = r \cup P_{new}((P_{new}(H) = r)) = 1)^c.$$

This event corresponds to proposition (10):

$$P_{new}(H) = r \text{ iff } P_{new}(P_{new}(H) = r) = 1,$$

which figures in Faithful Reflection and Confident Reflection. Clearly, one can extend these constructions to delineate nth-order credences for any n.

If we use a higher-order probability space $(\Omega, \mathfrak{F}, P, T)$ to model the agent's credences at t_1 regarding her credences at t_2, then we cannot simultaneously use it to model her credences at t_1 regarding her credences at other times. Luckily, Harsanyi's solution extends to many such higher-order credences. One need simply posit appropriate random variables $T_i: \Omega \to \Delta(\mathfrak{F})$. We can then use the expanded higher-order probability space $(\Omega, \mathfrak{F}, P, T_1, T_2, T_3, \ldots, T_n)$ to model the agent's credences at t_1 regarding her credences at t_1 (corresponding to T_1), her credences at t_1 regarding her credences at t_2 (corresponding to T_2), her credences at t_1 regarding her credences at t_3 (corresponding to T_3), and so on for other times t_i.

Spaces of the form $(\Omega, \mathfrak{F}, P, T_1, T_2, T_3, \ldots, T_n)$ are widely used in game theory to model higher-order credence. Mertens and Zamir (1985) prove that, if we begin with an ordinary measurable space $(\Sigma, \mathfrak{G})$ satisfying mild topological conditions, then we can embed $(\Sigma, \mathfrak{G})$ in a space $(\Omega, \mathfrak{F}, T_1, T_2, T_3, \ldots, T_n)$ rich enough to model all possible combinations of higher-order credence over $(\Sigma, \mathfrak{G})$. Heifetz and Samet (1998) give a somewhat different construction that makes no topological assumptions. See also (Meier, 2012). Game theorists typically use these formal constructions to model an agent's credences *regarding the credences of other agents* (Zamir, 2009). As Samet (1999) emphasizes, the constructions can just as well serve to model an agent's credences *regarding her own credences*.[11]

In what follows, I focus exclusively on credences that can be modeled using a higher-order probability space $(\Omega, \mathfrak{F}, P, T)$. As I have indicated, this setup already suffices to model numerous higher-order credences.

§7.3 Violations of Reflection

[11] Campbell-Moore (2015) formalizes higher-order probability using a fixed point construction based on Kripke's (1975) theory of truth. The formalism is rich enough to express certain kinds of *self-referential probability*, such as an interpreted sentence that describes the credence attached by the agent to that very sentence. See (Campbell-Moore, 2015) for details and for connections to the higher-order probability spaces used in game theory. And see (Caie, 2013) for discussion of how self-referential probability relates to Dutch book argumentation.

There are many options for $T: \Omega \to \Delta(\mathfrak{F})$, codifying different ways that the state of the world might determine an agent's credences at t_2. If we confine attention to Lewis learning scenarios, T must have a very simple form. In a Lewis learning scenario, the agent's new credences are given by an update rule $C: \mathfrak{E} \to \Delta(\mathfrak{F})$. Each update rule C determines a random variable $T: \Omega \to \Delta(\mathfrak{F})$ as follows:

(15) $T(\omega) = C_E$ *if* $\omega \in E$.

Call *T the random variable determined by C*. We can proceed as in §7.2, taking $(\Omega, \mathfrak{F}, P, T)$ to be the higher-order probability space. Note that the partition $\mathfrak{E}$ may include events that codify higher-order probabilities. Specifically, $\mathfrak{E}$ may include events of the form $P_{new}(H) = r$.

We can now analyze in more formal terms how Conditionalization and Reflection relate in Lewis learning scenarios. Let $\mathfrak{E}$ be a partition, and let C be the update rule defined by

(16) $C_E =_{df} P(. \mid E)$ *for all* $E \in \mathfrak{E}$.

Let T be the random variable determined by C. Assume that the partition $\mathfrak{E}$ contains $P_{new}(H) = r$. By (15) and (16),

$$T(\omega) = P(. \mid P_{new}(H) = r) \qquad \textit{if } \omega \in P_{new}(H) = r.$$

Substituting the official definition of $P_{new}(H) = r$, we have

$$T(\omega) = P(. \mid P_{new}(H) = r) \qquad \textit{if } \omega \in \{\omega: T(\omega)(H) = r\},$$

so that

(17) $T(\omega)(H) = P(H \mid P_{new}(H) = r)$ *if* $\omega \in \{\omega: T(\omega)(H) = r\}$

Pick an arbitrary ω such that $T(\omega)(H) = r$, and then (17) entails

(18) $P(H \mid P_{new}(H) = r) = r$.

Thus, an agent who conditionalizes in Lewis learning scenarios must conform to Reflection. This argument formalizes §5's observation that, in a Lewis learning scenario containing $P_{new}(H) = r$ as a partition proposition, non-reflectors are pushed onto the diagonal in Figure 2 and so are not conditionalizers. Given the factive Dutch book theorem, the following is an immediate consequence:

Dutch Book Theorem for Reflection: *Let* $(\Omega, \mathfrak{F}, P)$ *be a probability space, let* $\mathfrak{E}$ *be a partition, let C be an update rule for* $(\Omega, \mathfrak{F}, \mathfrak{E})$*, and let T be the random variable determined by C. Suppose that* $P(P_{new}(H) = r) > 0$*,* $P(H \mid P_{new}(H) = r) \neq r$*, and* $P_{new}(H) = r \in \mathfrak{E}$*. Then there exists a factive Dutch book for* $(\Omega, \mathfrak{F}, P)$ *and C.*

One could also prove the theorem directly, by formalizing van Fraassen's Dutch book.

No comparable theorem is possible for *generalized* Lewis learning scenarios, thanks to the converse non-factive Dutch book theorem. Suppose that P violates Reflection:

$$P(P_{new}(H) = r) > 0 \ \& \ P(H \mid P_{new}(H) = r) \neq r.$$

If C is the update rule defined by (16), then the converse non-factive Dutch book theorem entails that there does not exist a non-factive Dutch book for $(\Omega, \mathfrak{F}, P)$ and C. This result formalizes my claim from §5 that violations of Reflection do not by themselves induce vulnerability to a sure loss in generalized Lewis learning scenarios. There are possible agents who violate Reflection yet who are not Dutch bookable.[12]

[12] Suppose the agent becomes certain at t_2 of partition proposition E. A non-factive Dutch book must inflict a loss in outcomes lying outside $P_{new}(E) = 1$. Formally: (d*) requires $X + Y_E$ to inflict a loss in *all* outcomes, not just outcomes belonging to $P_{new}(E) = 1$. Why should we consider outcomes lying outside $P_{new}(E) = 1$? After all, such outcomes are not possible given that the agent is newly certain of E. My answer is that, *from the agent's point of view*, those outcomes are epistemically possible both at t_1 and t_2. At t_1, the agent may not know how she would react

Formally, the key difference between the factive and the non-factive case centers on (15). In a Lewis learning scenario *simpliciter*, the agent learns the true partition proposition E and forms new credences C_E. Thus, her new credences are modeled by the random variable determined by C via (15). In a generalized Lewis learning scenario, the true outcome ω may belong to partition proposition E even while the agent becomes certain of a different partition proposition E^*. Her new credences are then given by C_{E^*} rather than C_E, so that the random variable determined by C via (15) does not model her new credences. Once we abandon (15), the inference from (16) to (18) fails: it becomes possible for a conditionalizer to violate Reflection. To illustrate, consider the higher-order probability space (S, $\mathfrak{S}$, P, T) from §7.2. As we saw, this space violates Reflection. Consider a generalized Lewis learning scenario with partition

$$\mathfrak{E} = \{\{\omega_1\}, \{\omega_2\}\}$$

and update rule C defined by (16). Note that

$$T(\omega_1)(\{\omega_1\}) = 0 \neq 1 = P(\{\omega_1\} \mid \{\omega_1\}) = C_{\{\omega_1\}}(\{\omega_1\}),$$

so that (15) fails. This is a simple example where Reflection fails even though the update rule is Conditionalization. Much more complicated examples along similar lines are possible. When the update rule C is given by (16), the converse non-factive Dutch book theorem shows that there does not exist a non-factive Dutch book for (Ω, $\mathfrak{F}$, P) and C *whether or not P violates Reflection*. Similar reasoning applies to Adapted Reflection and Faithful Reflection.[13]

References

at t_2 to evidence E: she may not know that she would react by becoming certain of E (though that is in fact how she would react). At t_2, the agent may not know her own credence in E with complete certainty. Thus, a "sure loss" should inflict a loss even in outcomes lying outside $P_{new}(E) = 1$.

[13] I presented a draft of this paper at the Berkeley-Stanford Circle in Logic and Philosophy in Fall 2017. I thank all participants for their comments, especially Wesley Holliday, Thomas Icard, and Dana Scott. Thanks also to Richard Pettigrew, Sherrilyn Roush, Melanie Schoenberg, Snow Zhang, and a referee for this journal for helpful feedback.

Armstrong, D. M. (1963). Is introspective knowledge incorrigible?", *The Philosophical Review*, *72*, 417-432.

Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy*, 100, 356–70.

Briggs, R. (2009). Distorted reflection. *The Philosophical Review*, *118*, 59-85.

Caie, M. (2013). Rational probabilistic incoherence. *The Philosophical Review*, *122*, 527-575.

Campbell-Moore, C. (2015). How to express self-referential probability: a Kripkean proposal. *The Review of Symbolic Logic*, *8*, 680-704.

Chalmers, D. (2012). *Constructing the World*. Oxford: Oxford University Press.

Christensen, D. (2007). Epistemic self-respect. *Proceedings of the Aristotelian Society*, *107*, 319-337.

de Finetti, B. 1937/1980. Foresight. Its logical laws, its subjective sources. Rpt. in H. E. Kyburg Jr. & H. E. Smokler (Eds.) *Studies in Subjective Probability* (pp. 94-158). Huntington: Robert E. Krieger.

Earman, J. (1992). *Bayes or Bust?*. Cambridge: MIT Press.

Easwaran, K. (2011). Bayesianism I: introduction and arguments in favor. *Philosophy Compass*, *6*, 312-320.

Fristedt, B. & Gray, L. (1997). *A Modern Approach to Probability Theory*. Boston: Birkhäuser.

Gaifman, H. (1988). A theory of higher-order probabilities. In B. Skyrms & W. Harper (Eds.), *Causation, Chance, and Credence: Proceedings of the Irvine Conference and Probability and Causation* (pp. 191-219). Boston: Kluwer.

Hájek, A. (2009). Dutch book arguments. In P. Anand, P. Pattanaik & C. Puppe (Eds.), *The Handbook of Rationality and Social Choice* (pp. 173-195). Oxford: Oxford University Press.

Harsanyi, J. (1967). Games with incomplete information played by "Bayesian" players, I-III. Part I: the basic model. *Management Science*, *14*, 159–182.

Heifetz, A. & Samet, D. (1998). Topology-free typology of beliefs. *Journal of Economic Theory*, 82, 324-341.

Huber, F. (2016). Formal representations of belief. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016). plato.stanford.edu/archives/spr2016/entries/formal-belief/.

Huemer, M. (2007). Epistemic possibility. *Synthese*, *156*, 119-142.

Jeffrey, R. (1983). *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

---. (2004). *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.

Kolmogorov, A. N. (1933/1956). *Foundations of the Theory of Probability*. 2nd English ed., Trans. N. Morrison. New York: Chelsea.

Koons, F. (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge: Cambridge University Press.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, *72*, 690-716.

---. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Lasonen-Aarnio, M. (2015). I'm onto something! Learning about the world by learning what I think about it. *Analytic Philosophy*, *4*, 267-297.

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

Mahtani, A. (2012). Diachronic dutch book arguments. *The Philosophical Review*, *121*, 443-450.

---. (2015). Dutch books, coherence, and logical consistency. *Noûs*, *49*, 522-537.

Meacham, C. (2016). Understanding conditionalization. *Canadian Journal of Philosophy*, *45*,

767-797.
Meier, M. (2012). An infinitary probability logic for type spaces. *Israel Journal of Mathematics*, 192, 1-58.
Mertens, J.-F. & Zamir, S. (1985). Formulation of Bayesian analysis for games of incomplete information. *International Journal of Game Theory*, *14*, 1-29.
Pettigrew, R. (Forthcoming). Bayesian updating when what you learn might be false. *Erkenntnis*.
Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156-198). London: Routledge and Kegan.
Rescorla, M. (2018). A Dutch book theorem and converse Dutch book theorem for Kolmogorov Conditionalization. *The Review of Symbolic Logic*, *11*, 705-735.
---. (2021). On the proper formulation of Conditionalization. *Synthese*, *198*, 1935-1965.
---. (Forthcoming). An improved Dutch book theorem for Conditionalization. *Erkenntnis*.
Roush, S. (2009). Second-guessing: a self-help manual. *Episteme*, *6*, 251-268.
---. (2018). Knowledge of our own beliefs. *Philosophy and Phenomenological Research*, *96*, 45-69.
Samet, D. (1999). Bayesianism without learning. *Research in Economics*, *53*, 227-242.
Schwitzgebel, E. (2008). The unreliability of naïve introspection. *The Philosophical Review*, *117*, 245-273.
Skyrms, B. (1980). Higher-order degrees of belief. In D. H. Mellor (Ed.), *Prospects for Pragmatism: Essays in Honor of F.P. Ramsey* (pp. 109-137). Cambridge: Cambridge University Press.
---. 1987. Dynamic coherence and probability kinematics. *Philosophy of Science*, *54*, 1–20.
Sobel, J. (1987). Self-doubts and Dutch strategies. *Australasian Journal of Philosophy*, *65*, 56-81.
Talbott, W. (2015). Bayesian Epistemology. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. plato.stanford.edu/archives/sum2015/entries/epistemology-bayesian/.
Teller, P. (1973). Conditionalization and observation. *Synthese*, *26*, 218–258.
Uchii, S. (1973). Higher order probabilities and coherence. *Philosophy of Science*, *40*, 373-381.
van Fraassen, B. (1984). Belief and the will. *The Journal of Philosophy*, *81*, 235–56.
---. (1995). Belief and the problem of Ulysses and the sirens. *Philosophical Studies*, *77*, 7–37.
---. (1999). Conditionalization, a new argument for. *Topoi*, *18*, 93-96.
Vickers, J. (2000). I believe it, but soon I'll not believe it any more: skepticism, empiricism, and reflection. *Synthese*, *124*, 155-174.
Vineberg, S. (2011). Dutch book arguments. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2011). plato.stanford.edu/archives/sum2011/entries/dutch-book/.
Weisberg, J. (2007). Conditionalization, Reflection, and self-knowledge. *Philosophical Studies*, *135*, 179-197.
---. (2009). Varieties of Bayesianism. In D. Gabbay, S. Hartman & J. Woods (Eds.), *Handbook of the History of Logic*, vol. 10 (pp. 477-551). New York: Elsevier.
Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
Zamir, S. (2009). Bayesian games: games with incomplete information. In R. Meyers (Ed.), *Encyclopedia of Complexity and System Sciences* (pp. 426-441). New York: Springer.

Outcome	***Net Payoff***
H & $P_{new}(H) = r$	$s - 1$
$\neg H$ & $P_{new}(H) = r$	s
$P_{new}(H) \neq r$	0

Table 1. Net payoff for the conditional bet offered at t_1.

Outcome	***Net payoff***
$P_{new}(H) = r$	$(r - s)[1 - P(P_{new}(H) = r)]$
$P_{new}(H) \neq r$	$(s - r)[P(P_{new}(H) = r)]$

Table 2. Net payoff for the sidebet offered at t_1.

Outcome	***Net payoff***
H	$1 - r$
$\neg H$	$-r$

Table 3. Net payoff for the bet offered at t_2.

Outcome	***Net Payoff***
H & $P_{new}(H) = r$	$(s - r)[P(P_{new}(H) = r)]$
$\neg H$ & $P_{new}(H) = r$	$(s - r)[P(P_{new}(H) = r)]$
$P_{new}(H) \neq r$	$(s - r)[P(P_{new}(H) = r)]$

Table 4. Net payoff for the bets from Tables 1, 2, and 3, computed under the assumption that the third bet is offered iff $P_{new}(H) = r$.

Outcome	*Net payoff*
$P(H) = r$	1
$P(H) \neq r$	– 1

Table 5. Since the agent is certain that $P(H) = r$, expected value is 1.

Outcome	*Net payoff*
Heads	1
Tails	– 1

Table 6. A bet on a coin toss. Expected value is 0.

Outcome	*Net Payoff*
H & $P(H) = r$	$s - 1$
$\neg H$ & $P(H) = r$	s
$P(H) \neq r$	0

Table 7. Net payoff for the conditional bet on H.

Outcome	*Net payoff*
$P(H) = r$	$(r - s)[1 - P(P(H) = r)]$
$P(H) \neq r$	$(s - r)[P(P(H) = r)]$

Table 8. Net payoff for the sidebet on $P(H) = r$.

Outcome	*Net payoff*
H	$1 - r$
$\neg H$	$- r$

Table 9. Net payoff for the bet on H offered when $P(H) = r$.

Outcome	***Net Payoff***
H & $P(H) = r$	$(s - r)[P(P(H) = r)]$
$\neg H$ & $P(H) = r$	$(s - r)[P(P(H) = r)]$
$P(H) \neq r$	$(s - r)[P(P(H) = r)]$

Table 10. Net payoff for the bets from Tables 7, 8, and 9, computed under the assumption that the third bet is offered iff $P(H) = r$.

Outcome	***Net payoff***
H & E	$1 - P(H \mid E)$
$\neg H$ & E	$- P(H \mid E)$
$\neg E$	0

Table 11. Net payoff for the conditional bet offered at t_1.

Outcome	***Net payoff***
E	$[P(H \mid E) - C_E(H)][1 - P(E)]$
$\neg E$	$- [P(H \mid E) - C_E(H)]P(E)$

Table 12. Sidebet on E offered at t_1.

Outcome	***Net payoff***
H & E	$C_E(H) - 1$
$\neg H$ & E	$C_E(H)$
$\neg E$	0

Table 13. Net payoff for the conditional bet offered at t_2.

Outcome	***Net payoff***
H & E	$P(E)[C_E(H) - P(H \mid E)]$
$\neg H$ & E	$P(E)[C_E(H) - P(H \mid E)]$
$\neg E$	$P(E)[C_E(H) - P(H \mid E)]$

Table 14. Net payoff for the bets from Tables 11, 12, and 13.

$\boldsymbol{H \,\&\, P_{new}(H) = r}$	$(s-1)+(r-s)[1-P(P_{new}(H)=r)]$	$(s-r)[P(P_{new}(H)=r)]$
$\boldsymbol{\neg H \,\&\, P_{new}(H) = r}$	$s+(r-s)[1-P(P_{new}(H)=r)]$	$(s-r)[P(P_{new}(H)=r)]$
$\boldsymbol{H \,\&\, P_{new}(H) \neq r}$	$(s-r)[P(P_{new}(H)=r)]$	$(s-r)[P(P_{new}(H)=r)]+1-r$
$\boldsymbol{\neg H \,\&\, P_{new}(H) \neq r}$	$(s-r)[P(P_{new}(H)=r)]$	$(s-r)[P(P_{new}(H)=r)]-r$
	$\boldsymbol{P_{new}(P_{new}(H) = r) \neq 1}$	$\boldsymbol{P_{new}(P_{new}(H) = r) = 1}$

Table 15. Net payoffs when the non-reflector faces van Fraassen's bookie under the assumption that (10) is true. Rows reflects possible outcomes for the bets offered at t_1 and t_2. Columns reflect whether the agent and bookie become certain of $P_{new}(H) = r$ at t_2. In the first column, the bookie does not become certain of $P_{new}(H) = r$ and so does not offer the bet from Table 3. In the second column, the bookie becomes certain of $P_{new}(H) = r$ and so offers the bet from Table 3; since we have assumed (10), it follows that $P_{new}(H) = r$ and hence that the non-reflector accepts the bet. The table lists payoffs for accepted bets in all possible outcomes. However, only the grey outcomes are consistent with (10).

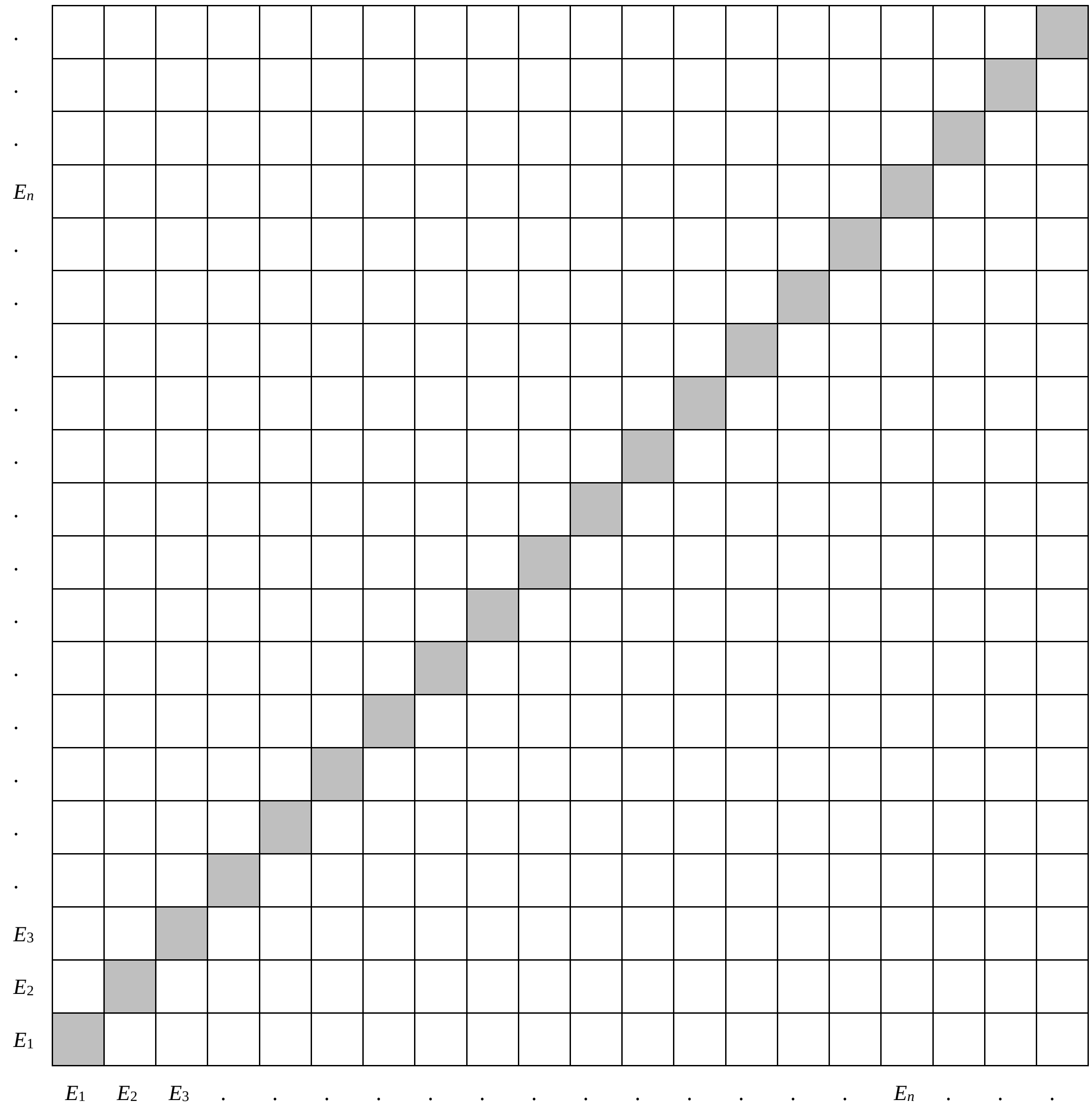

Figure 1. The horizontal axis depicts the partition proposition of which the agent becomes certain at t_2. The vertical axis depicts the true partition proposition. Thus, the horizontal axis determines which bet is enacted at t_2. In a Lewis learning scenario, only cells along the diagonal are possible. In a generalized Lewis learning scenario, all cells are possible. A non-factive Dutch book ensures a net loss at every cell. A factive Dutch book only ensures a net loss along the diagonal. (The partition $\mathcal{E}$ may be countably infinite, in which case Figure 1 only depicts a finite portion of the infinitely many possible outcomes.) Reprinted with permission from Springer Nature Customer Service Center GmbH: Springer Nature, *Erkenntnis*, "An Improved Dutch Book Theorem for Conditionalization" (Rescorla, forthcoming).

$\boldsymbol{P_{new}(H) = r}$		
$\boldsymbol{P_{new}(H) \neq r}$		
	$\boldsymbol{P_{new}\ (P_{new}(H) = r) \neq 1}$	$\boldsymbol{P_{new}\ (P_{new}(H) = r) = 1}$

Figure 2. Rows reflect the agent's first-order credence in H at t_2. Columns reflect whether she is certain that $P_{new}(H) = r$ at t_2. In a Lewis learning scenario that contains $P_{new}(H) = r$ as a partition proposition, she must fall in one of the grey cells. In a generalized Lewis learning scenario, she may fall in one of the white cells.