Dutch Book and Accuracy Theorems[1]

Author: Anna Mahtani
Affiliation: London School of Economics and Political Science
Address: 24 Station Road, Wokingham, Berkshire, RG40 2AE
Email: a.mahtani@lse.ac.uk

Abstract

Dutch book and accuracy arguments are used to justify certain rationality constraints on credence functions. Underlying these dutch book and accuracy arguments are associated theorems, and I show that the interpretation of these theorems can vary along a range of dimensions. Given that the theorems can be interpreted in a variety of different ways, what is the status of the associated arguments? I consider three possibilities: we could *aggregate* the results of the differently interpreted theorems in some way and motivate rationality constraints based on this aggregation; we could be *permissive* and accept the conclusions of the dutch book and accuracy arguments under all interpretations of the associated theorems; or we could *select* one uniquely correct interpretation of the dutch book/accuracy theorem and use that to justify certain rationality constraints. I show that each possibility faces problems, and conclude that dutch book and accuracy theorems cannot be used to justify any principle of rationality.

I Introduction

On the orthodox Bayesian view, we can model a rational agent's epistemic state with a probability function assigning numbers to each claim that the agent can entertain.[2] A probability function conforms to the probability axioms, and here I give one version of these axioms.

We let {E, F, E_1, … F_n} be the set of claims to which the credence function assigns values, and set out the axioms as follows:

(1) 0≤P(E) for any E
(2) If E is a tautology, then P(E)=1

[1] Many thanks to Vasiliy Romanovskiy who read an earlier draft of this paper and gave me some invaluable comments; many thanks too to the audience at the Proceedings of the Aristotelian Society for the very thought-provoking questions on this work.

[2] Most theorists would claim that an agent may have a credence in a claim that (s)he is not currently entertaining, but many would resist the claim that agents have credences in all possible claims. Questions over the range of claims in which an agent has credences are explored in the literature on awareness growth (Bradley 2017, Karni and Vierø 2013, Mahtani Forthcoming, Steele and Stefánsson Forthcoming).

(3) If E and F are incompatible, then $P(E \vee F) = P(E)+P(F)$.

As I have set out the probability axioms, they are rather vague. What exactly is a claim? Is it a sentence, a proposition, or something else? What does it mean for a proposition to be a tautology? And under what circumstances are two claims incompatible? In response to these questions an orthodox Bayesian might explain that there is an underlying set of *states*. A claim (or 'event') corresponds to a set of these states; a tautology is the set of all states; and two claims are incompatible if they have no states in common. Our interest then shifts to the question: what are these states? Are they possible worlds? Impossible worlds? Or what? There are different ways of answering these questions, and so different versions of probabilism. I discuss some of these different versions in this paper, but in this introduction, I leave these matters undecided.

Dutch book and accuracy arguments have been given for these axioms: dutch book arguments by (Ramsey 1931, De Finetti 1964), and accuracy arguments by (Joyce 1998). I will give the flavour of these arguments in the next section by setting out dutch book and accuracy arguments for axiom (2): similar arguments for the other axioms can be given. The dutch book and accuracy arguments each involve an associated theorem, and in section III I show that there are many different ways of interpreting these theorems. Given that there are many different interpretations of these theorems, what is the status of the dutch book and accuracy arguments which aim to justify certain rationality constraints on our credences, most prominently that they should conform to the probability axioms described above? I address this question in section IV, concluding that the dutch book and accuracy arguments cannot be used to justify any principle of rationality.

II Dutch book and accuracy arguments

I begin by giving examples of dutch book and accuracy arguments. I focus on arguments for axiom (2), but (as mentioned above) similar arguments for the other axioms can also be given.

Take an agent with credence function Cr that violates axiom (2): that is, for some tautology, E, Cr(E) is $\underline{v}$, and $\underline{v}$ does not equal 1. Let us suppose here that $\underline{v}$ is less than 1: a similar argument could be given were we to suppose instead that $\underline{v}$ is more than 1. Given that Cr(E) is $\underline{v}$, we could offer this agent the following bet, and (s)he would consider it fair:[3] the agent gets £v and pays back £1 iff E is true.[4] But given that E is a tautology, it is guaranteed to be true and therefore the agent is guaranteed to make a loss, for (s)he gets £$\underline{v}$ but then inevitably

[3] In Christensen's terms, the agent's epistemic attitude would sanction the bet as fair (Christensen 1996).

[4] One (albeit imperfect — see (Rabinowicz 2014)) way to see that the agent would consider the bet fair is to calculate the agent's expected utility for accepting the bet (assuming here that the agent values only money, and values that linearly). The agent has a credence of $\underline{v}$ that E is true, and so (s)he has a credence of $\underline{v}$ that (s)he will end up with £($\underline{v}$-1); the agent (let us assume) has a credence of 1-$\underline{v}$ that E is false, and so (s)he has a credence of 1-$\underline{v}$ that (s)he will end up with £$\underline{v}$. Thus the expected utility of accepting the bet is $£(\underline{v})(\underline{v}-1) + £(1-\underline{v})(\underline{v}) = \underline{v}^2-\underline{v}+\underline{v}-\underline{v}^2 = 0$. Thus the agent is indifferent between accepting and rejecting the bet.

has to pay out £1, and given that £$\underline{v}$ is less than £1 the agent ends up with less money than (s)he had before. Because the agent would consider fair a set of bets that is guaranteed to result in a loss — that is, a 'dutch book' of bets — we say that the agent is dutch bookable, and the idea is that we infer from this that the agent is irrational.[5] There is a debate in the literature over how we can justify this move from the claim that the agent is dutch bookable to the conclusion that the agent is irrational. Some think that we have only shown that the agent has a practical defect (which is that in certain — perhaps very unlikely — scenarios the agent is guaranteed to lose money), and that we cannot move from the claim that an agent has a practical defect to the conclusion that the agent has an epistemic defect (Hájek 2008). Some think that the best version of the argument bypasses any claims about pragmatic error and directly demonstrates epistemic incoherence (Skyrms 1980, Christensen 1991, Christensen 1996, Howson and Urbach 1993). To side-step this controversy for now, we can separate dutch book *arguments* from dutch book *theorems*. A dutch book theorem states that an agent with a particular sort of credence function is dutch bookable — that is, will accept as fair a bet or set of bets that is guaranteed to result in a loss. Here then is an example of a dutch book theorem that we seem to have established above: an agent who violates probability axiom (2) is dutch bookable. We could give arguments for further dutch book theorems and thereby show that any agent who violates any of the probability axioms is dutch bookable. A dutch book argument takes the further step of claiming that dutch bookable agents are irrational, and so that any agent who violates any of the probability axioms is irrational.[6] I will discuss dutch book arguments again later in the paper, but my initial focus is on interpreting dutch book theorems.

Having given the flavour of a dutch book argument and theorem, I turn now to the accuracy argument and theorem. The accuracy of an agent's credence in a claim depends on how big the difference is between the agent's credence and the truth-value of the claim. If a claim is true then it has truth-value 1, and the further the agent's credence is from 1 the less accurate it is; conversely if a claim is false then it has truth-value 0, and the further the agent's credence is from 0, the less accurate it is. There are various ways of measuring the distance between an agent's credence function and the truth value, and a number of criteria have been proposed that any acceptable measure should meet (Joyce 1998). One popular measure that satisfies these criteria is the Brier score, and using this score we can measure the inaccuracy of an agent's credence function as follows: for each claim that is assigned a credence, we take the difference between the truth-value of that claim and the credence value assigned, and square that difference; we then sum these squared differences for all the relevant claims to arrive at measure of the inaccuracy of the credence function as a whole.

Let us now consider again an agent with credence function Cr that violates axiom (2) — so for some tautology E, Cr(E) is $\underline{v}$ where $\underline{v}\neq 1$. We contrast Cr with an alternative credence function Cr', which is similar to Cr in all respects except that it assigns a value of 1 to E.

[5] In the paper I write as though either an agent or a credence function can be dutch bookable: an agent is dutch bookable iff his or her credence function is.

[6] Perhaps to justify this further step we would need both a dutch book theorem and a converse dutch book theorem — that is, to establish that an agent whose credence function has a particular feature is irrational, we need to show both that any agent whose credence function has such a feature is dutch bookable, and that any agent whose credence function does not have such a feature is not dutch bookable (Hájek 2005). For simplicity I set this complication to one side.

Which of these credence functions is the most accurate? In measuring the inaccuracy of each using the Brier score, the only different component in the total will be the square of the difference between the truth-value of E (which must be 1 as E is a tautology) and the value assigned to E by the credence function: for Cr, this gives $(1-\underline{v})^2$ which is greater than 0, whereas for Cr' this is gives $(1-1)^2$ which is 0. Thus Cr' has a lower inaccuracy score and so is more accurate than Cr. What is more, Cr' is guaranteed to be more accurate than Cr, because E is a tautology and so is guaranteed to have a truth-value of 1. Because Cr' is guaranteed to be more accurate than Cr, we say that Cr' *accuracy dominates* Cr.[7] The idea is that from the fact that Cr is accuracy dominated we infer that an agent with credence function Cr is irrational. More generally, we can show that any credence function that violates any of the probability axioms is accuracy dominated, and so argue that any agent with such a credence function is irrational.

But we might question why it is that an agent with a credence function that is accuracy dominated is irrational. One reason to question this is that there may be other virtues that compete with accuracy, and it may be that the agent's accuracy dominated credence function has some of these virtues while the accuracy-dominating credence function does not.[8] Here we can sidestep this interesting issue by making a move parallel to that made for the dutch book argument above. We can distinguish between an accuracy argument and an accuracy theorem. An accuracy theorem is a claim that any credence function with a certain feature is accuracy dominated: an accuracy argument takes the further step of claiming that an agent with an accuracy dominated credence function is irrational. My initial focus is on the interpretation of the accuracy theorem.

I have now given a quick overview of dutch book and accuracy arguments and theorems. But these overviews have been somewhat vague, for they involve the expression 'guarantee', and this term has not been sharply defined. What does it mean, exactly, for a set of bets to be *guaranteed* to result in a loss? And what is it for one credence function to be *guaranteed* to be more accurate than another? The vagueness in this term 'guarantee' is transmitted to the expressions 'dutch bookable', and 'accuracy dominated', and so to the concept of dutch book and accuracy theorems. In this paper, I explore a range of different ways of specifying the terms that have been left vague, and this gives us a range of interpretations of the dutch book and accuracy theorems.

III Interpretations of the theorems

The dutch book theorem involves the claim that some set of bets is guaranteed to make a loss. Analogously, the accuracy theorem involves the claim that some set of credence functions are guaranteed to be accuracy dominated. Below I explore a range of ways of interpreting these theorems. I focus on the dutch book theorem, but all that I say can also be applied to the accuracy theorem. I begin in III.i with a standard and intuitive interpretation (the 'base-interpretation'), and then in III.ii - III.iv I turn to alternatives.

[7] In the paper I write as though either an agent or a credence function can be accuracy dominated: an agent is accuracy dominated iff his or her credence function is.

[8] For further discussion on this point, see for example (Easwaran and Fitelson 2012, Pettigrew 2013).

III.i The base-interpretation

We can say that a set of bets is guaranteed to make a loss if and only if that set of bets would make a loss when assessed at each possible world. It can be proved that any credence function that violates any of the probability axioms is dutch bookable in this sense — provided that we choose the right reading of the probability axioms, with states corresponding to possible worlds.

This is probably the most intuitive reading of the dutch book theorem and associated concepts, and we can take it as our base reading in what follows. But I pause here to note that this base reading itself fractures into multiple interpretations, for there are many ways to understand the idea of 'possible worlds'. We might take possible worlds to be metaphysically possible worlds, centred possible worlds, epistemically possible worlds, or even impossible worlds — though whether we can make sense of all of these ideas is debateable (Jackson 2011, Bjerring 2013). Each of these different interpretations of 'possible world' gives rise to a different reading of the probability axioms. For example, if we take possible worlds to be *metaphysically* possible worlds, then any metaphysically necessary statement — such as that George Orwell is Eric Blair — will count as a tautology.[9] And the dutch book theorem will rule that any agent who has a credence of less than 1 in this claim is dutch bookable, because this agent will accept as fair a bet that will lose her money at every metaphysically possible world. If on the other hand we take possible worlds to be *epistemically* possible worlds, then perhaps the claim that George Orwell is Eric Blair will not count as a tautology: this will depend on how exactly we spell out the nature of epistemically possible worlds, and possibly on the agent's state of knowledge.

Let us turn now from this base-interpretation — though noting that it is really a set of interpretations — to a first sort of alternative interpretation.

III.ii Restricting the set of states

On the base-interpretation, a set of bets is *guaranteed* to make a loss if and only if those bets would make a loss at each possible world. But on an alternative interpretation, the worlds in which the bets are required to make a loss can be smaller than the set of all worlds. Here is one such interpretation which has played an important role in the literature on dutch book arguments. On this view, a book of bets is guaranteed to make a loss if and only if it makes a loss at every possible world in which the agent has his or her actual credence function.[10] Thus we assess the book of bets against what is true at each possible world at which the agent has his or her actual credence function: if the set of bets make a loss at every such world, then on

[9] 'George Orwell' and 'Eric Blair' are both proper names designating the same object, and so (at least on standard accounts of proper names) they designate the same object in every possible world where that object exists. To be precise, the relevant necessary truth here not that George Orwell is Eric Blair, but rather the conditional that if George Orwell exists, then he is Eric Blair. I skate over this for simplicity.

[10] As for the first interpretation, here we also could consider a range of interpretations of 'possible world'.

this reading the agent is dutch bookable. This interpretation is often implicitly assumed[11], sometimes explicitly assumed (Milne 1991) and at least once explicitly rejected (Briggs 2009).

To see the appeal of this reading, consider the imaginary bookie often summoned up to introduce the dutch book argument. We imagine this bookie having information about the target agent's credence function and using this knowledge to design a set of bets. If the bookie can design a set of bets that he or she is certain will lose the agent money — using nothing more than his or her knowledge of the agent's credence function — then we think that the agent's credence function must be somehow defective. This train of thought suggests that the relevant possible worlds at which we must assess the outcome of the bet are those possible worlds in which the agent has his or her actual credence function — and this is obviously a smaller set than the set of all possible words.

Agents who are dutch bookable on the base reading are also dutch bookable on this reading, for if a set of bets loses money at every possible world, then it will certainly lose money at every possible world at which the agent has his or her actual credence function. But in addition on this reading we also class as dutch bookable any agent who lacks perfect knowledge of his or her own credence function, for such an agent will accept as fair a bet on his or her own credence function which is guaranteed (in the relevant sense) to make a loss: that is, it will make a loss at every world at which the agent has his or her actual credence function.[12]

III.iii Changing the nature of the states

Another way in which we can depart from the base-reading is by interpreting the states as something other than possible -worlds.[13] One such alternative interpretation (Mahtani 2015), draws on a standard definition of logical consistency (Halbach 2010). This definition works with sentences rather than propositions, and with 'structures' rather than possible worlds. A structure assigns meaning to every non-logical part of the language. For example, at a given structure, 'George Orwell' means George Orwell, and 'author' means author; at another structure, 'George Orwell' means Barack Obama, and 'author' means surgeon. Thus a sentence such as 'George Orwell is an author' can be true under one structure but false under another, depending on what meanings each structure assigns to the relevant terms involved. Logical terms such as 'and', 'not', and 'all' have meanings that do not vary across structures. Because of this, some sentences — the tautologies — are true at all structures, for they are

[11] It is by implicitly assuming this reading of dutch book arguments that we can get a dutch book argument for (synchronic) reflection (van Fraassen 1984).

[12] To see this, take an agent with a credence function Cr, where Cr assigns 0.5 to some claim P but assigns $\underline{v}$ (where $\underline{v}$ is less than 1) to the claim that Cr assigns 0.5 to P. Then the bookie can offer the following bet, which the agent will accept as fair: the agent gets £$\underline{v}$, but pays back £1 iff (s)he does have a credence of 0.5 in P. At every possible world where the agent has his or her actual credence function, this bet will result in a loss, and so on this reading the agent is dutch bookable.

[13] It is debateable how broadly the term 'possible world' should be understood. There may be readings on which the states described in this section fall under the umbrella of 'possible worlds', in which case this is a mere variant on the base-reading covered in 3.1.

true in virtue of their logical form; similarly some sentences — contradictions — are false at all structures.[14]

Applying this to the Bayesian framework, we can see a credence function as an assignment of values to sentences rather than propositions. On this framework, the bets that a credence function will endorse as fair will be bets waged over sentences rather than over propositions. Some sets of bets waged over sentences are guaranteed to result in a loss in the following sense: they lose money no matter what meanings are assigned to the non-logical terms of the relevant sentences — in other words, they lose money at every structure. We can say that a credence function that endorses such a set of bets is dutch-bookable, and the credence functions that are dutch-bookable in this sense are exactly those that violate the probability axioms — provided that we read the probability axioms as part of a framework on which the states are structures and the 'events' are sentences.

III.iv Changing the assessment

A final way in which the interpretation of the dutch book theorem can vary is in how the bets are assessed at each state. So far we have been assuming that the bets are assessed against what is true at each state, but other options are possible: for example, we can assess the bets against what is _verified_ rather than true at each state.

This alternative interpretation of the dutch book argument is not discussed in the literature as far as I know, but it is a natural development of two-dimensionalism. On Chalmers' two-dimensionalist account (Chalmers 2011), an assertion has both a primary and a secondary intension, and here I focus on the primary intension, which is (roughly) the set of centered (metaphysically) possible worlds which _verify_ the assertion. A centred possible world verifies an assertion iff under the hypothesis that you inhabit that centred possible world[15], your ideal credence[16] in the assertion is 1. To explore this idea, consider a non-actual world _w_* in which one man is christened and known as 'Eric Blair' and never becomes an author, while some other man calls himself 'George Orwell' and writes _Animal Farm_, _1984_ and so on under this pen-name. Under the hypothesis that you inhabit this world _w_* — that this is how things actually are — your ideal credence in the assertion that George Orwell is not Eric Blair is (let's say) 1. Thus this is a world that verifies the assertion that George Orwell is not Eric Blair. Notice here that there is a difference between what is true at _w_* and what _w_* verifies. For while _w_*verifies the assertion that George Orwell is not Eric Blair, it is not the case that the assertion is true at _w_*.[17] After all it is actually the case that George Orwell is Eric Blair, and furthermore this is metaphysically necessary and so true at all metaphysically possible

[14] For an introduction to this classical account of the semantics of predicate logic, see (Halbach 2010).

[15] It makes a difference how this hypothesis is stated — that is, how the centred possible world is described. For Chalmers' purposes, the relevant description is given in what he calls a 'canonical specification' of a world, which is (roughly) a complete description of the world given only in neutral vocabulary, where neutral vocabulary excludes proper names such as 'George Orwell' and 'Eric Blair'.

[16] By this Chalmers means the 'credences that the subject should have on ideal rational reflection' (Chalmers 2011, 621)

[17] The assertion is made in the actual world, so the sentence means what it actually does, and we are evaluating the truth of this assertion at other possible worlds.

worlds, and furthermore at all centred possible worlds.[18] Thus the assertion that George Orwell is not Eric Blair is not true at *w**, even though it is verified there.

This suggests an alternative interpretation of the dutch book theorem: we can say that a bet is guaranteed to result in a loss iff it results in a loss at each centered possible world when assessed against what is verified (rather than true) at that world. The credence functions that are dutch bookable in this sense are those that violate the probability axioms — provided that we understand these axioms in a parallel way, with a tautology defined as a claim that is verified at each centered possible world, and so on.

III.v Theoretical interpretations

We have seen that we can vary the interpretation of the dutch book theorem along several dimensions: we can impose various restrictions on which objects belong in our set of states; we can vary the nature of the states — treating them as possible worlds of various sorts, or interpretations, or something else; and we can vary the way in which we assess a bet at a given state. I have described several examples of interpretations drawn from the literature, each with its own rationale. But many other interpretations are possible, as theoretically there is a vast number of alternative options along each of these three dimensions, and there may also be other dimensions that I have not mentioned along which the interpretation can vary. Thus besides the specific examples of interpretations mentioned, there are many — perhaps infinitely many — other ways that the dutch book and accuracy theorems could be interpreted.[19] And of course all that I have said in this section about dutch book theorems can also be said about accuracy theorems.

In the next section, I turn to question the status of the dutch book and accuracy arguments, given that the associated theorems are open to interpretation as we have seen.

IV. Rationality

The interest in the dutch book and accuracy theorems is that they form part of an argument for certain rationality constraints. I discuss this below in relation to the dutch book argument, but as usual all that I say can also be applied to accuracy arguments. We extend the dutch book theorem into a dutch book argument by drawing a link between dutch bookability and irrationality: the key claim is that any agent whose credence function is dutch bookable is thereby shown to be irrational. As we have seen, there are multiple interpretations of the dutch book theorem, and which credence functions are classed as dutch bookable depends on the interpretation. What rationality constraints can we then infer?

[18] As before, strictly speaking, what is metaphysically necessary is that if George Orwell exists then he is Eric Blair. I am simplifying here but the main point is unaffected.

[19] One way to see that the number of interpretations may be infinite is to consider the range of ways in which the set of states could theoretically be restricted. If we take the states to be possible worlds, and assume that there are an infinite number of possible worlds, then there will also be an infinite number of subsets of those possible worlds.

Below I set out some of the ways that we might respond. Firstly I consider two ways in which we might *aggregate* the results of the dutch book theorems under their various interpretations. Secondly, I consider the *permissive* option of endorsing the move from dutch bookability to irrationality under every interpretation of the dutch book theorem, coining a wide range of senses of 'irrationality' in the process. Thirdly and finally I consider the option of *selecting* one interpretation of the dutch book theorem as uniquely correct. I find problems with each of these options, and conclude that the dutch book and accuracy arguments cannot be used to argue for or justify claims about rationality.

I begin by considering two ways that we might aggregate the interpretations.

IV.i Aggregation

We have a variety of ways of interpreting the dutch book theorem, and corresponding to each is the set of possible agents whose epistemic state can be represented by a credence function that is dutch bookable under that interpretation. For example, if we consider the base interpretation, with possible worlds understood as metaphysically possible worlds, then the set of dutch bookable agents will include those who have a credence of less than 1 in any metaphysically necessary truth (such as that George Orwell is Eric Blair); and each alternative to this interpretation will carve out its own set of dutch bookable agents. Thus we are left not with a single set of dutch bookable agents, but rather with multiple such sets corresponding to the various interpretations of the dutch book theorem. How then can we move from dutch bookability to irrationality? How can we infer from all these various sets of dutch bookable agents that some particular associated set of agents are irrational? One approach is to try to somehow aggregate the sets of agents that are dutch bookable under these various interpretations.

Here I consider just two ways of aggregating. The first way I consider is to take the union: that is, the set of all agents who are classed as dutch bookable under *any* interpretation of the dutch book theorem. The second way is to take the intersection: that is, the set of all agents who are classed as dutch bookable under *every* interpretation of the dutch book theorem. The idea in either case would be to take the resulting set (whether the union or the intersection), and argue that the agents in this set have thereby been shown to be irrational.

Let us begin with the union of the sets — that is, the set of all agents who are dutch bookable under *any* interpretation of the dutch book theorem. It seems that this set would contain all — or nearly all — possible agents. For consider that besides the interpretations of the dutch book theorem which are mentioned in the literature, there are many other interpretations that are theoretically possible. For example, there are interpretations under which a set of bets is guaranteed to result in a loss iff it results in a loss at any state in some *restricted* set. If we assume that the states in question are possible worlds, then we can, theoretically, restrict the relevant set to just the actual world — or perhaps even to the empty set. On this reading, any set of bets that would lose money in the actual world (if we restrict the relevant set to just the actual world), or any set of bets at all (if the relevant set is the empty set) would count as *guaranteed* to make a loss, for it would make a loss at all states in the relevant set. Thus we have readings of the dutch book theorem on which all credence functions are classed as dutch bookable unless they assign 1 to all truths and 0 to all falsehoods — and arguably readings on which every possible credence function is classed as dutch bookable. Thus at least all non-omniscient agents — and perhaps even all possible agents — would count as dutch bookable on this proposal. The union of the sets of agents who are dutch bookable under *some* interpretation would thus comprise all or nearly all possible agents, and so this view would force us to conclude that all or nearly all possible agents are irrational, thereby trivialising the concept of rationality.

A natural response here is to object that a reading on which the set of states is restricted to just the actual world, or indeed to no worlds at all, is not a plausible reading of the dutch book theorem. Some readings of the dutch book theorem are merely theoretically possible readings with no rationale, while others have some rationale behind them and have been defended in the literature. Perhaps we should be aggregating just those sets of agents that are dutch bookable according to some reading that has a rationale and is defended in the literature? But this seems too arbitrary: whether a reading of the dutch book argument makes an appearance in the literature can depend on all sorts of extraneous factors, and whether there is *some* rationale for a given reading can depend on irrelevant features of the context. Surely what we want here is not the set of readings which have a rationale or have appeared in the literature, but rather the set of readings — or perhaps the unique single reading — that *truly* justifies the step from dutch bookability to irrationality. This will require selecting from amongst the possible readings, and I turn to this option in 4.3.

Before moving on, let us briefly consider an alternative way of aggregating, and this is to take the intersection rather than the union of the sets of agents that are classed as dutch bookable under the various possible readings. This looks no more promising: if we consider all theoretically possible readings, then the intersection will be empty, for there will be two interpretations that class disjoint sets of agents as dutch bookable; and as before, an attempt to limit the interpretations to those that appear in the literature would be arbitrary.

IV.ii Permissiveness

An alternative possibility is simply to accept the move from dutch bookability to irrationality under every interpretation of dutch bookability, but with the term ‘irrationality’ permitting a range of different interpretations. For example, take the base reading of the dutch book theorem, with possible worlds understood as metaphysically possible worlds. On this reading, a certain set of agents are classified as dutch bookable — and this set includes for example agents who have a credence of less than 1 that Eric Blair is George Orwell. We can make the move from dutch bookability to irrationality and claim that this set of agents is indeed irrational, but here we should understand that ‘irrational’ has a special reading which does not match up perfectly with our everyday sense of irrational (for in the everyday sense, of course, an agent who has a credence of less than 1 that Eric Blair is George Orwell is not classed as irrational). Now we can consider a different reading — for example the base-reading with possible worlds understood as epistemically rather than metaphysically possible worlds. On this reading, a different set of agents will be classified as dutch bookable, and the idea is that we can also extend this dutch book theorem into a dutch book argument, and claim that this set of agents is irrational, but here ‘irrational’ has another special reading. We can continue in this way for all the theoretically possible readings of the dutch book argument, and so have multiple different sets of possible agents, each classed as irrational, but with ‘irrational’ understood differently for each set.

This idea may be coherent, and Bayesian Epistemologists are generally open to the idea that the term 'rational' can be given a technical sense that differs from its everyday sense. But this proliferation in the senses of the term 'rational' drains the term of its meaning. Perhaps many of the categories carved out by the various senses of 'dutch bookable' are of interest, but it is not at all clear what it would add to class them as 'irrational' in any sense. It is better and less misleading simply to recognise that the term 'dutch bookable' has many readings, each carving out a certain set of credence functions, and that some of these categories have little to do with what we mean by the term 'irrationality'. The move from dutch bookability to irrationality is not justified across the board.

IV.iii Selection

Here we come to what may seem like the obvious response, which is to say that though there are many ways of reading the dutch book theorem, only one of them authorizes the move from dutch bookability to irrationality.[20] Which reading is it?

We might try to give a formal reason for privileging one reading, and a natural move here is to claim that the narrowest reading is to be preferred — that is, the reading that leads to the smallest set of possible agents being classed as dutch bookable. It is questionable whether this is a well-motivated move, but there are in any case two other more practical problems with this option. Firstly, it may be that the narrowest reading leads to no possible agents being classed as dutch bookable: whether we can make sense of some such reading will depend on what sorts of objects we are willing to admit might be classed as possible states, and what sorts of different ways of assessing bets we would consider. Secondly, there may be more than one equally narrow reading. For example, if we take the narrowest reading on which bets are assessed against what is ***true*** at any possible world, it seems that there will be an equally narrow reading (catching a similarly-sized but different set of credence functions) on which bets are assessed against what is ***false*** at any possible world. There seems to be no good purely formal reason to choose one of these readings over the other.

Without a formal way of choosing between the interpretations, what considerations can we draw on in order to choose which interpretation warrants the move from dutch bookability to irrationality? The various interpretations of dutch bookability carve out various sets of agents: which of those sets are thereby classed as irrational? One natural move here is to examine these sets, and see which could most plausibly be described as irrational.[21,22] Intuitively, an

[20] Another related option is to claim that there is some limited set of privileged readings, and then apply one of the approaches considered under 4.1 or 4.2 to that set. This option suffers from the same defect as the more straightforward option described in this section.

[21] When we do so, we may find that none of these sets maps perfectly onto that set that we might have pre-theoretically labelled as irrational, for the term 'irrational', as used in its everyday context, has multiple strands of meaning, many of which have no echo in the relatively precise classifications that we get from each reading of the dutch book and accuracy theorems. But we may not be drawing just on our pre-theoretic use of the term 'irrational', for our use of the term has been trained by theory.

[22] Here I am assuming that we look just at the extensions of these sets — that is we look simply at which agents each sets contain. And so long as we focus just on the sets' extensions, we will judge two interpretations of dutch bookability to be on a par if they end up classifying the very same set of

agent who has a credence of less than 1 in some metaphysically necessary truth is not thereby irrational, and this is seen as a reason to reject the base reading with ‘possible world’ interpreted as ‘metaphysically possible world’. Perhaps some other reading of ‘possible world’ gives us a better result? We might try taking ‘possible world’ to mean some sort of epistemically possible world, honing this concept to give us the most plausible reading. Or we might try varying the way in which bets are assessed at worlds, perhaps using Chalmer’s idea of verification rather than truth to help us get the result that we seek. In these various manoeuvres, we are using our judgment about which agents are rational to select the right interpretation of the dutch book theorem — that is, the interpretation of the dutch book theorem under which the move from dutch bookability to irrationality gives us plausible results. Given this, what is the status of the dutch book argument?

I claim that the dutch book argument cannot be used for the purpose for which it was designed. It cannot be used to justify the claim that any particular set of agents (for example agents whose credence functions violate certain principles) are irrational, for if anyone doubts the claim then it is open to them to simply challenge the interpretation of the dutch book theorem: why should the dutch book theorem be interpreted in that particular way, so that that particular set of agents ends up classed as irrational? The choice of interpretation itself requires justification. To defend the choice of interpretation by pointing out that it classes as dutch bookable only agents who could reasonably be classed as irrational, will not be an adequate answer. To select the interpretation with an eye on the desired conclusion is to beg the question against someone who does *not* agree with the conclusion. This is why I claim that the dutch book argument cannot be used justify classing any particular set of agents as irrational, and so it cannot be used to justify probabilism, conditionalization, or any other principle.[23] And as usual, what holds for the dutch book argument holds for the accuracy argument too.

V Conclusion

Dutch-book and accuracy arguments have been criticised in various ways in the literature. Here I have put forward a further reason to object to these arguments. I have shown that there are a variety of ways of interpreting the dutch book and accuracy theorems, and this leaves

agents as dutch bookable. An alternative is to also consider the intensions of the sets: is there a way of selecting the correct interpretation on this basis? We can understand one set to contain all those agents who are guaranteed to lose money *as a matter of metaphysical necessity*; another set to contain all those agents whose loss if guaranteed *a priori*; other sets will contain agents whose loss is guaranteed *on the basis of logic* (for various different systems of logic); and so on. But even once the intensions of the sets are taken into account, there doesn’t seem to be any obvious reason to select one of these sets in particular — except that it contains those agents that we already consider to be irrational. Many thanks to Guy Longworth and Matt Parker for pressing me on these points.

[23] My claim here is that we cannot use dutch book arguments to *justify* any rational restrictions. It does not follow that dutch book arguments can play no role at all in the framework. Perhaps dutch bookability and rationality can be inter-definable — as analogously some theorists see knowledge and safety as inter-definable (Williamson 2000) — and we should not then expect an independently justified characterisation of dutch bookability. But there will then be no direct argument from the claim that an agent is dutch bookable to the claim that that agent is irrational: it will always be open to an interlocutor to challenge the interpretation of dutch bookability. Many thanks to Rory Madden for this suggestion.

the defenders of the associated arguments with a trilemma. Firstly, the defenders might attempt to aggregate the sets of agents classed as dutch bookable/accuracy dominated under these interpretations, but under the two most obvious ways of aggregating these sets, we are left with a trivial classification. Secondly, the defenders might endorse the arguments under every interpretation of the relevant theorems, coining multiple senses of the term ‘irrational’, but these multiple senses drain the term of content. Thirdly and finally, the defenders might select some particular interpretation of the dutch book/accuracy theorem as privileged, and claim that under this interpretation alone the theorem can be extended into an argument, but I have argued that there is no non-question-begging reason for selecting one particular interpretation over another. For these reasons, I claim that dutch book and accuracy theorems cannot be used to justify any principle of rationality.

References

Bjerring, Jens. 2013. ‘Impossible worlds and logical omniscience: an impossibility result.’ *Synthese* 190 (13): 2505-2524.

Bradley, R. 2017. *Decision Theory with a Human Face*. Cambridge: CUP.

Briggs, Rachael. 2009. ‘Distorted Reflection.’ *Philosophical Review* 118 (1): 59-85.

Chalmers, David J. 2011. ‘Frege's Puzzle and the Objects of Credence.’ *Mind* 120 (479): 587-635.

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

Christensen, David. 1996. ‘Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers.’ *Journal of Philosophy* 93: 450-479.

De Finetti, Bruno. 1964. ‘Foresight: Its Logical Laws, Its Subjective Sources.’ In *Studies in Subjective Probability*, by H E Kyburg and H E Smokler, 93-158. New York: Wiley.

Easwaran, Kenny, and Brandon Fitelson. 2012. ‘An ‘Evidentialist’ Worry About Joyce’s Argument for Probabilism.’ *Dialectica* 66 (3): 425-433.

Hájek, Alan. 2008. ‘Dutch Book Arguments.’ In *The Oxford Handbook of Rational and Social Choice*, by Paul Anand, Prasanta Pattanaik and Clemens Puppe. Oxford: OUP.

Hájek, Alan. 2005. ‘Scotching Dutch Books?’ *Philosophical Perspectives* 19 (issue on Epistemology), ed. John Hawthorne. (1): 139-151.

Halbach, Volker. 2010. *The Logic Manual*. Oxford: OUP.

Howson, Colin, and Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court.

Jackson, Frank. 2011. ‘Possibilities for Representation and Credence: Two Space-ism versus One Space-ism.’ In *Epistemic Modality*, by Andy Egan and Brian Weatherson. Oxford: OUP.

Joyce, James M. 1998. ‘A Nonpragmatic Vindication of Probabilism.’ *Philosophy of Science* 65 (4): 575-603.

Karni, E, and M L Vierø. 2013. ‘"Reverse Bayesianism": A choice-based theory of growing awareness.’ *American Economic Review* 103 (7): 2790-2810.

Mahtani, Anna. Forthcoming. ‘Awareness growth and dispositional attitudes.’ *Synthese*

Milne, Peter. 1991. ‘A dilemma for subjective bayesians — and how to resolve it.’ *Philosophical Studies* 62 (3): 307 - 314.

Pettigrew, Richard. 2013. ‘Accuracy and Evidence.’ *Dialectica* 67 (4): 579-596.

Rabinowicz, Wlodek. 2014. ‘Safeguards of a Disunified Mind.’ *Inquiry* 57 (3): 356-383.

Ramsey, Frank P. 1931. ‘Truth and Probability.’ In *The Foundations of Mathematics and other Logical Essays*, by Frank P Ramsey, 156-198. Oxford: Routledge.

Skyrms, Brian. 1980. ‘Higher Order Degrees of Belief.’ In *Prospects for Pragmatism: Essays in Honor of F. P. Ramsey*, by D H Mellor. Cambridge: CUP.

Steele, Katie, and Orri Stefánsson. Forthcoming. *Severe Uncertainty: Reasoning with Unknown Possibilities*. Cambridge: CUP.

van Fraassen, Bas C. 1984. ‘Belief and the Will.’ *Journal of Philosophy* 81 (5): 235-256.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: OUP.