NOÛS

NOÛS 49:3 (2015) 522–537
doi: 10.1111/nous.12070

Dutch Books, Coherence, and Logical Consistency[1]

ANNA MAHTANI
London School of Economics

Abstract

In this paper I present a new way of understanding Dutch Book Arguments: the idea is that an agent is shown to be incoherent iff (s)he would accept as fair a set of bets that would result in a loss under any interpretation of the claims involved. This draws on a standard definition of logical inconsistency. On this new understanding, the Dutch Book Arguments for the probability axioms go through, but the Dutch Book Argument for Reflection fails. The question of whether we have a Dutch Book Argument for Conditionalization is left open.

1. Introduction

Dutch Book Arguments (DBAs) have given us some results that we want, and some results that we don't. On the up-side we have DBAs that show that coherent agents have credence functions that obey the probability axioms, and a DBA to show that coherent agents conditionalize. On the down-side, we have a DBA that seems to show that coherent agents have perfect access to their own credence functions, and a DBA that seems to show that coherent agents always obey the implausible Reflection Principle. Unless we can stomach these unappealing results, it seems that we must reject all DBAs, and so cannot use them to motivate the results that we want. In this paper, I argue for a new way of understanding DBAs. On this new understanding we get to keep the DBAs that we want, and reject those that we don't.

I begin (in section 2) by discussing synchronic DBAs, which have been used to show that coherent agents have credence functions that obey the probability axioms (a good result), and that coherent agents have perfect access to their own credence functions (a bad result). I explain (in section 3) how, on my new understanding, just the right synchronic DBAs go through. I then turn (in section 4) to diachronic DBAs, and explain how (on the old-understanding of DBAs) we have a DBA for the Reflection Principle—an unwelcome result. I show that this DBA does not go through on my new understanding of DBAs. Then (in section 5) I contrast my understanding of DBAs with that of Rachael Briggs (2009). Finally (in section 6) I consider whether the DBA for conditionalization goes through, and conclude that my new understanding of DBAs leaves room for maneuver on this issue.

2. Synchronic DBAs

I begin with a synchronic Dutch Book Argument. Alan has a credence of 0.6 in the claim that all whales are mammals (W), and a credence of 0.5 in the claim that not all whales are mammals (¬W). A bookie could offer Alan the following two bets:

Bet A	
W	-£0.60 + £1 = £0.40
¬W	-£0.60

Bet B	
W	-£0.50
¬W	-£0.50+£1 = £0.50

For bet A, Alan pays out £0.60, and gets £1 back iff W; for bet B, Alan pays out £0.50, and gets £1 back iff ¬W. Given Alan's credence function, he would consider both of these bets to be fair. Yet they are certain to jointly result in a loss for him, for he pays out a total of £1.10, and whatever happens—whether W obtains or not—Alan will get back exactly £1.00. Thus the bets are guaranteed to lose him £0.10—and we say that Alan has been 'Dutch Booked'. This is supposed to show that Alan is incoherent.

We can generalize this argument to show that for any (determinate) claim ψ and any value v, if an agent has a credence of v in ψ and some credence other than $(1 - v)$ in $\neg\psi$, then a Dutch Book like the one above can be made against the agent. Thus (the argument runs) any such agent must be incoherent. In fact, we can produce DBAs to demonstrate that any coherent agent will obey all of the probability axioms. This is a good result. The probability axioms seem sensible enough, but we don't just have to rely on our intuition to justify our acceptance of them: here we have an *argument* to the conclusion that any agent who violates them is incoherent. But can we rely on DBAs—or do they sometimes lead us astray?

To answer this question, I focus first on the move in the DBA above where we said that the bets were *guaranteed* to lose Alan money. This is an important move, because clearly a perfectly coherent agent can accept a set of bets that *as it happens* will lose him or her money. For example, suppose that an agent (Betty) is certain that a fair coin has just been tossed, though she cannot see the result. Betty accepts as fair a bet (C) in which she pays out £0.50, and gets back £1 iff the coin landed heads. Now suppose that in fact the coin landed tails, and Betty loses money. This does not show that Betty is incoherent, because though bet C loses her money, it was not *guaranteed* to lose her money. But what does 'guaranteed' mean here exactly? In what way was Alan guaranteed to lose money? Obviously it is not necessary that Alan will lose money on bets A and B—for there are possible worlds where he has a different credence function and will not accept them as fair. The idea is rather that it is necessary that *if* Alan accepts the bets as fair, then he will lose money on them.

The underlying thought is that whether an agent is coherent (with respect to his or her credence function) depends *just* on that agent's credence function. It does not depend on how the rest of the world is. Thus we might imagine holding the agent's credence function fixed, and so holding fixed the agent's assessment as fair of some particular book of bets, and varying the rest of the world. If the book of

bets that the agent would accept as fair always results in a loss for the agent—no matter how the rest of the world varies—then the agent has been shown to be incoherent.

To put the point vividly, we might imagine a bookie who has perfect access to the agent's credence function. We can imagine the bookie looking at what I will call a 'credence spreadsheet' for the agent, which has a list of claims in one column (all the claims that the agent has some credence in) and in the next column the values that the agent's credence function assigns to each claim. The bookie has no other information about how the world is, because whether the agent is incoherent does not depend on how the rest of the world is: it depends just on the agent's credence function. From the information in the credence spreadsheet, the bookie tries to design a book of bets that he knows that the agent will accept as fair, and that he knows will lose the agent money. If he is able to do this, then the agent can be Dutch Booked, and so is shown to be incoherent. This way of understanding how DBAs work is what I will call the 'old way', and it seems to be the understanding that Milne is working with (Milne 1991). Milne states that a book of bets does not count as a Dutch book if the bookie 'has not guaranteed profit on all possible outcomes, just on the actual one', and clarifies that because of the information that the bookie has about the agent's credence function (or, as Milne writes 'degree of belief') 'when he [the bookie] sets the stakes there are no longer open possibilities in which the proposition concerning her [the agent's] degree of belief is false, if actually true, or true, if actually false.' (Milne 1991: 308).

On this old understanding of how DBAs work, they lead us astray. Here is a simple example to illustrate the point.[2] Charlotte's credence in the claim that London is a capital city (L) is exactly 0.75 ($Cr(L) = 0.75$), but Charlotte is not certain of this fact about her own credence. Let's say that Charlotte has a credence of 0.8 that $Cr(L) = 0.75$. It seems that Charlotte may nevertheless be perfectly coherent. To be coherent, an agent isn't required to be certain of every true claim—and that seems to include true claims about herself. An agent can be coherent without being certain what his or her blood group is, or whether (s)he is in love, and it seems that similarly she can be coherent without being certain of every true claim about her own credence function. The problem is that it seems that Charlotte can be Dutch-Booked. For a bookie can offer her the following bet:

Bet D

$Cr(L) = 0.75$	£0.80-£1.00 = -£0.20
$Cr(L) \neq 0.75$	£0.80

Because Charlotte's credence in ($Cr(L) = 0.75$) is 0.8, she will accept bet D, which will give her a loss of £0.20. And we can say that she is *guaranteed* to make a loss on this bet, because facts about her own credence function determine not just that she will accept the bet, but what the outcome of that bet will be. Imagine the bookie looking at Charlotte's credence spreadsheet: the bookie can tell just from

the information available to him about the agent's credence function both that she will accept the bet, and that she will lose money on it.

It seems then that—on the old way of understanding DBAs—Charlotte has been Dutch Booked, and is classed as incoherent. More generally, any agent who lacks perfect access to his or her own credence function is classed as incoherent. This would be an unwelcome result. It may be tempting to think that there should be *some* sort of fit between a coherent agent's credence function and the credence function that (s)he thinks (s)he has, but it is certainly excessive to require absolute certainty. Thus DBAs—understood in the old way—lead us astray here. In the next section, I explain how on my new understanding of Dutch Book Arguments, this problem does not arise.

3. Interpretations

To motivate my account, I begin by thinking about outright beliefs rather than credence functions. What is it for an agent's belief state to be coherent? I think that the simplest and best answer here is that an agent's belief state is coherent iff the set of all the claims that the agent believes form a logically consistent set.

What is it for a set of claims to be logically consistent? Here I take a standard line (Halbach 2010): a set of claims is logically consistent iff there is an interpretation under which those claims are all true. An 'interpretation' will assign meanings to all the non-logical terms in the language. So the sentence 'All pencils are fish', under some interpretation means that all cats are mammals—which is true. Thus though a person who believes that all pencils are fish is obviously deluded, he is not thereby incoherent, for the content of his belief is true under some interpretation. In contrast, take a person who believes both 'All pencils are fish' and 'there is a pencil that is not a fish'. There is no interpretation under which these two claims are both true, and so the two claims are logically inconsistent, and this person is in an incoherent belief state.

This seems like a clear and compelling definition of incoherence with respect to outright beliefs. How can we adapt it to give us a definition of incoherence with respect to credence functions? The key idea was that to assess whether a person has a coherent outright belief state, we take the set of claims believed, and vary the *interpretation* of those claims: iff under every interpretation there is some claim in the set that is false, the agent is incoherent. Similarly, then, to assess whether a person has a coherent credence function, we take that agent's credence function—and some book of bets that she would accept as fair—and vary the *interpretation* of the relevant claims: iff under every interpretation the agent makes a loss, then it follows that the agent is incoherent.

To put the point vividly, imagine again our bookie who is viewing his credence-spreadsheet for an agent. We can lift the requirement that the bookie knows nothing about how the world is, other than facts about the agent's credence function: bookies are now allowed to know other facts about how things are in the world. The new constraint is that the bookie does not know how to interpret the claims in the first column of the spreadsheet: he is sure of the meaning of the logical terms, and he understands the structure of the sentences, but he does not know what the subject

specific terms mean. Thus for example if a claim in the first column is 'All whales are mammals', then the bookie does not know whether this means that all whales are mammals, or that all fish are pencils. We might equivalently imagine that the first column contains formalizations of all the claims, with the dictionary hidden from the bookie. Let us call this a 'credence spreadsheet (logical form version)'. The bookie then comes up with a book of bets, which he knows the agent will accept as fair. These will be 'written in the same language' as the claims in the credence spreadsheet (logical form version). So for example, if the bookie can see that the agent has a credence of 0.6 in the claim 'All whales are mammals', then he can include a bet at the relevant rate on the claim 'All whales are mammals'—and know that the agent will accept it as fair. Whether the bet results in a profit or a loss for the agent will depend on the interpretation of the claim 'All whales are mammals'. I claim that the agent is shown to be incoherent only if some book of bets that the agent accepts as fair will lose the agent money *under any interpretation* of the claims in that book of bets.

With this new understanding of Dutch Book Arguments in mind, let us return to our cases of Alan and Charlotte. Alan is the agent who has a credence of 0.6 in claim W (All whales are mammals) and a credence of 0.5 that in claim ¬W (Not all whales are mammals). The bookie has access to this information, but does not know how to interpret the claims—i.e. he does not know whether 'whales' means whales, or fish, etc. As before, the bookie offers him the following two bets, which Alan will accept as fair:

Bet A

W	-£0.60 + £1 = £0.40
¬W	-£0.60

Bet B

W	-£0.50
¬W	-£0.50+£1 = £0.50

No matter what sentence W means, these two bets will result in a loss for Alan. If W means that all whales are mammals, then W is true, in which case Alan will lose £0.10. On the other hand, if W means that all fish are pencils, then W is false, in which case Alan will lose £0.10. Under any interpretation, these two bets result in a loss. Thus Alan is classed as incoherent. In fact (though I don't show it here) on my understanding of how Dutch Book Arguments work, every agent who violates the probability axioms is classed as incoherent. This is a good result.

Now let us compare the case of Charlotte. She has a credence of 0.8 in claim L (that London is large), and a credence of 0.75 in the claim that her credence in L is 0.8. Let's suppose again that the bookie offers Charlotte the following bet:

Bet D

Cr(L) = 0.75	£0.80-£1.00 = -£0.20
¬Cr(L) = 0.75	£0.80

Will bet D lose her money under any interpretation? It is not obvious what the logical form of 'Charlotte's credence in L is 0.75' (i.e. 'Cr(L) = 0.75') is. Perhaps

the logical form of this sentence is just Pa (in which case, for all the bookie knows, the sentence means that David Cameron is a horse). Or perhaps the logical form is Pab—or perhaps it has some other more complex logical form. In any case, we can focus on an interpretation under which all the terms have their actual meanings, except for the term 'credence' which means 'half-credence' which we define as follows: for any claim ϕ and value v, an agent has a half-credence of v in ϕ iff she has a credence of $2v$ in ϕ. Under this interpretation, the sentence means that Charlotte's half-credence in L is 0.8—and so the sentence is false.[3] Thus there is an interpretation under which bet D will result in a profit for Charlotte rather than a loss, and so (on my new understanding of DBAs) Charlotte has not been shown to be incoherent. This is a good result, because intuitively an agent like Charlotte who lacks perfect access to her own credence function may nevertheless be coherent.

Having described my new understanding of DBAs, and shown how it works in synchronic cases, I turn now to Diachronic DBAs.

4. Diachronic DBAs

So far we have been concerned just with 'synchronic coherence'—i.e. the coherence of an agent *at* a time. I turn now to the issue of 'diachronic coherence'—i.e. the coherence of an agent *across* time. These are the two diachronic coherence principles that I discuss:

> Conditionalization: Take an agent with a credence function Cr. Take any claim E such that $Cr(E) > 0$. Take any claim P, such that Cr assigns (P&E) a value. We can define $Cr(P/E)$ as $Cr(P\&E)/Cr(E)$. The principle of conditionalization then states that if this agent learns just E, and nothing else, then if the agent is coherent, his or her new credence function Cr_E will be such that $Cr_E(P) = Cr(P/E)$.
>
> Reflection: Take an agent with credence function Cr_0 at time t_0, and consider some future time t_1, when the agent will have credence function Cr_1. Take any claim P and any value v, such that $Cr_0(Cr_1(P) = v) > 0$. The Reflection Principle states that unless $Cr_0(P/Cr_1(P) = v) = v$, this agent is incoherent.

Diachronic DBAs have been offered for both Conditionalization (Lewis 1999) and Reflection (Van Fraassen 1984). It is clear that Reflection places unreasonable demands on an agent: an agent who merely suspects that she might forget something, or that she might misinterpret future evidence—and so does not automatically defer to her future credence function—is (intuitively) not thereby incoherent (Christensen 1991: 234–235, Talbott 1991: 138–140, Briggs 2009: 64–66). Thus it might seem that we must reject all diachronic DBAs to avoid being saddled with the counterintuitive Reflection Principle. This would mean that we could not use a diachronic DBA to argue for conditionalization—or for any other diachronic coherence principle. Fortunately, on my new understanding of how DBAs work, the DBA for Reflection fails. And it fails for reasons that have nothing to do with the fact that it is diachronic: a parallel DBA for a synchronic version of the principle fails too. Thus we have a decisive reason to reject the DBA for the Reflection Principle, that

leaves open the option of accepting a DBA for Conditionalization—or some other diachronic principle of coherence.

I begin with an example of an agent, Delia, who violates Reflection. Delia has read that kebabs are healthy (H) but is not quite convinced. At the start of the evening, her credence that kebabs are healthy is 0.7 (i.e. $Cr(H) = 0.7$). She suspects though that she might get drunk later, and by 10pm she might be irrationally convinced (with a credence of 0.9) that kebabs are healthy. She currently has a credence of 0.2 that by 10pm she will have a credence of 0.9 that kebabs are healthy (i.e. $Cr(Cr_{10pm}(H) = 0.9) = 0.2$). But even under the supposition that by 10pm her credence in H will be 0.9, her current credence in H is still 0.7 (i.e. $Cr(H/Cr_{10pm}(H) = 0.9) = 0.7$). This agent can be dutch-booked, using the following 3 bets:

Bet E, to be offered at the start of the evening

$Cr_{10pm}(H) = 0.9$	-£0.04 + £0.20 = £0.16
$Cr_{10pm}(H) \neq 0.9$	-£0.04

Bet F, offered at the start of the evening

$Cr_{10pm}(H) = 0.9$ & H	£0.70-£1 = -£0.30
$Cr_{10pm}(H) = 0.9$ &¬H	£0.70
$Cr_{10pm}(H) \neq 0.9$	£0

Bet G, offered at 10pm iff $Cr_{10pm}(H)=0.9$

H	-£0.90 + £1 = £0.10
¬H	-£0.90

We can see that Delia would accept each of these bets if offered, but is certain to lose £0.04 overall. For either $Cr_{10pm}(H) \neq 0.9$, in which case bets F and G are both either not offered or called off, and Delia loses £0.04 on bet E; or $Cr_{10pm}(H) = 0.9$, in which case Delia gains £0.16 on bet E, but loses £0.20 on bets F and G together, resulting in an overall loss of £0.04. Thus it seems that Delia has been Dutch Booked.

Before I assess how this agent fares given my new understanding of Dutch Book Arguments, I pause here to consider more generally how this sort of diachronic Dutch Book Argument is supposed to work. It is clear that the bookie cannot implement his strategy if the only information he gets at all is information about Delia's credence function at the *start* of the evening—for then how will he know at 10pm whether or not to offer bet G? We cannot allow the bookie to have a 'strategy' of offering a bet iff some particular state of affairs obtains unless the bookie *knows* (or will know at the appropriate time) whether or not this particular state of affairs obtains, and so is able to implement his strategy. To see this, consider again our agent Betty who is certain that a fair coin has been tossed, but has not seen the result. A bookie might have a strategy of offering her bet B (in which she pays out £0.50 and gets £1 back iff the coin has landed heads) iff the coin has landed tails. This 'strategy'—if the bookie could implement it—would result in a sure loss for Betty. But Betty is not incoherent, and we will not allow this sort of betting 'strategy'. Presumably we allow the bookie the strategy of offering Delia bet G iff her credence in H at 10pm is 0.9, because we are imagining that the bookie will be able to use his information about Delia's credence function at 10pm to implement his strategy.

Thus the bookie needs to have information not just about Delia's credence function at the start of the evening, but also about Delia's credence function at 10pm. Should we imagine, then, that the bookie has information about the agent's credence function across all time? We could imagine the bookie looking at a sort of multidimensional graph, with time along one axis, and the agent's credence in each claim marked along some dimension. For any claim and any time, the bookie can look at the graph to find out the agent's credence in that claim at that time. But this picture is clearly not what we want. A bookie with access to *this* sort of information about an agent's credence function would be able to siphon money from any agent whose credence function changes in any way across time. For example, take an agent who starts the morning with a credence of ½ in the claim that the cricket will be cancelled. By lunchtime, his or her credence in the claim has increased to ¾. This agent may be perfectly coherent: perhaps dark storm clouds gathered mid-morning. But if a bookie could know from the start of the morning how this agent's credence function would develop, then the bookie could offer the agent one bet (H) at the start of the morning (the agent gets £0.50, and pays back £1 iff the cricket is cancelled), and another bet (I) at lunchtime (the agent pays £0.75, and gets £1 back iff the cricket is cancelled), resulting in a sure loss for this agent of £0.25. Bookies should not be able to Dutch book agents so easily—so we must drop the idea that the bookie has complete access to the agent's credence function across all time. A better idea is to imagine that the bookie gets information in 'real-time'. At any time, the bookie knows what the agent's current credence function is—but he has no special access to information about what the agent's credence function will be in the future.[4] We might imagine then that the bookie has a credence spreadsheet that he can 'refresh' at any time.

On the old understanding, the bookie gets to see the agent's credence spreadsheet, and we are now supposing that he can refresh it in real time. Thus in our example above involving Delia, the bookie will be able to plan and implement his strategy, and will know from the start of the evening that Delia will accept all bets offered, and that they will result in a loss. Thus—on the old understanding of DBAs—Delia has been shown to be incoherent; more generally, any agent who violates the Reflection Principle can be shown to be incoherent. This is a bad result, because intuitively coherent agents can violate the Reflection Principle.

Let's now consider whether this DBA works on my new understanding. We suppose again that the bookie has access to the refreshable credence spreadsheet, but that he does not know how to interpret the claims in the first column. Thus he knows that Delia has a credence of 0.7 in some claim H which has the logical form of 'kebabs are healthy', but he doesn't know what this means. He can also figure out that she has a credence of 0.7 in this claim (whatever it means) under the supposition of some other claim which has the logical form of '$Cr_{10pm}(H) = 0.9$', but he doesn't know what this claim means either. For all he knows, '$Cr_{10pm}(H) = 0.9$' means that Delia's half-credence at 10pm in H is 0.9. At 10pm, the bookie gets to see whether Delia's credence in 'H' (whatever 'H' means) has increased to 0.9. It seems then that the bookie is able to carry out his strategy. He can offer bets E and F at the start of the evening, and then at 10pm he can refresh his spreadsheet,

and offer bet G iff he sees that the agent has a credence of 0.9 in 'H' (whatever 'H' means). The problem is that he cannot be sure that this strategy will result in a loss for the agent, for under some interpretations, his strategy will give the agent a profit.

To see this, suppose first that Delia's credence in 'H' at 10pm is 0.9. Then all 3 bets will be offered. Take an interpretation under which '$Cr_{10pm}(H) = 0.9$' is false (e.g. take the interpretation under which it means that Delia's *half*-credence in H is 0.9), and 'H' is true. Under this interpretation, Delia will lose £0.04 on bet E, bet F will be called off, and Delia will make £0.10 on bet G, resulting in an overall profit for Delia of £0.06. Now suppose instead Delia's credence in 'H' at 10pm is still 0.7. Then just bets E and F will be offered. Take an interpretation under which '$Cr_{10pm}(H) = 0.9$' is true (e.g. take an interpretation under which it means that Delia's one-and-two-sevenths-credence in H = 0.9), and under which 'H' is false. Under this interpretation, Delia will win £0.16 on bet E, and win another £0.70 on bet F, resulting in an overall profit of £0.86. Thus clearly the bookie's strategy will not lose Delia money under every interpretation, and so (on my new understanding of DBAs), Delia is not shown to be incoherent.

We can understand why the argument fails here. To make a Dutch Book against Delia, the bookie relies on the assumption that either 1) bets F and G will both not be in force, and the agent will lose money on bet E, or 2) bets F and G will both be in force, and so will jointly result in a loss for the agent, which will more than outweigh the profit she makes on bet E. On my new understanding of Dutch Book Arguments, this assumption no longer holds. The bookie can ensure that bet G is in force iff Delia's credence at 10pm in 'H' (whatever 'H' means) is 0.9—for his strategy is to offer G under just these circumstances. But he cannot ensure that bet F is in force iff Delia's credence at 10pm in 'H' is 0.9. For bet F is a conditional bet, and whether it is in force depends on whether '$Cr_{10pm}(H) = 0.9$' is true—which in turn depends on how this claim is interpreted. If Delia's credence at 10pm in 'H' is 0.9, then there will nevertheless be interpretations under which '$Cr_{10pm}(H) = 0.9$' is false—and so there will be interpretations under which bet F is called off even though bet G is in force; similarly, if Delia's credence at 10pm in 'H' is not 0.9, then there will nevertheless be interpretations under which '$Cr_{10pm}(H) = 0.9$' is true—and so there will be interpretations under which bet F is in force even though bet G is not offered. Thus the bookie cannot be sure that bets F and G will either be in force or not in force together—and so cannot know that his strategy will result in a loss for Delia.

Thus on my new understanding of DBAs, the DBA for Reflection clearly fails. It fails because some of the bets involved are bets about the agent's own credence function—and the bookie's strategy depends on these bets being interpreted in a particular way (e.g. in such a way that '$Cr_{10pm}(H) = 0.9$' is true iff Delia's credence at 10pm in 'H' is 0.9). The failure of the DBA for the Reflection principle has nothing to do with the fact that the Reflection Principle and associated DBA are diachronic. In fact, a parallel synchronic DBA for a synchronic version of the Reflection Principle (according to which a coherent agent defers to his own current credence function) would also fail. This is a good result, because the synchronic

Reflection Principle faces counterexamples.[5] Thus on my new understanding of DBAs, we have decisive reason to reject the DBA for the Reflection Principle.

As I shall show, we do not have the same decisive reason to reject the DBA for Conditionalization. I begin with an agent Fred, who violates Conditionalization. Fred is a science student who has thought up an exciting but unlikely hypothesis, H. He thinks that E would offer a lot of support to this hypothesis—but hardly conclusive support: $Cr(H/E) = 0.7$. The student runs an experiment to test whether E. At the end of the experiment, let's suppose that the student will have either learnt just E and nothing else, or he will have learnt that ¬E. If he learns that E, then he will get wildly excited and overestimate the support that E gives to his hypothesis: $Cr_E(H) = 0.9$. We suppose that he thinks that E is fairly unlikely: $Cr(E) = 0.2$. This agent can be dutch-booked, using the following 3 bets:

Bet J, to be offered before the experiment runs

E	-£0.04 + £0.20 = £0.16
¬E	-£0.04

Bet K, offered before the experiment runs

E&H	£0.70-£1 = -£0.30
E&¬H	£0.70
¬E	£0

Bet L, offered after the experiment has run, iff the agent learns that E

H	-£0.90 + £1 = £0.10
¬H	-£0.90

To get this Dutch Book for Conditionalization to work—on either the old or new understanding of how Dutch Books work—we have to supplement the information that the bookie gets about the agent's credence function. Before the experiment, the bookie needs to know not just what value Cr assigns to various claims, but what value Cr_E assigns to various claims. He needs to know this *before* the experiment, because otherwise he could not design his strategy and know that it will succeed. Recall our case of the coherent agent who (on seeing dark storm clouds) updated her credence that the cricket would be cancelled: there are bets that a bookie could offer this agent at the start of the morning and at lunchtime, that the agent would accept as fair, and that would jointly result in a loss for him or her—but that should not count as a Dutch Book, or we will end up classing as incoherent every agent who changes her credence function. We must have higher standards for a Dutch Book: for an agent to be Dutch Booked, the bookie must be able to design a strategy that is risk-free—that he *knows* will lose the agent money.

The DBA for Conditionalization can only work then, if we suppose then that the credence spreadsheet that the bookie has access to is more complex than previously supposed. Not only does it state, for every claim ϕ that the agent's current credence function (Cr) assigns a value to, the value assigned to that claim (i.e. $Cr(\phi)$): it also states, for every two claims ψ and ϕ that Cr assigns a value to, the value Cr_ϕ assigns to ψ. Whether this is a reasonable addition to the credence spreadsheet is an issue that I take up in section 6. For now, I just note that something like this is required on the old understanding of DBAs—if the DBA for Conditionalization is to go through. And for the new understanding of DBAs, we make the same

adjustment—except as usual we add that the bookie does not know how to *interpret* the claims (ϕ, ψ and so on).

With this clarified, we can now consider whether the DBA for Conditionalization works on my new understanding of DBAs. It is clear that even if the bookie only has the logical form version of the (new, more complex) spreadsheet, he is nevertheless able to carry out his strategy and know that it will succeed. The bookie offers bets J and K before the experiment runs; then after the experiment he refreshes his credence-spreadsheet, and offers bet G iff he sees that Fred has a credence of 1 in 'E' (whatever 'E' means). The bookie can be sure that all bets offered will be accepted. The bookie can also be sure that Fred will lose money on these bets. Either 'E' (whatever it means) is false—in which case bets K and L will both either not be offered or be called off, and the agent will lose money on bet J; or 'E' (whatever it means) is true, in which case all 3 bets will be offered, and Fred will gain £0.16 on bet J but lose £0.20 on bets K and L together. Fred loses money on these bets *under every interpretation*. Thus Fred—and any agent who violates conditionalization—has been shown to be incoherent.

Thus on my new way of understanding DBAs, there is decisive reason to reject the DBA for Reflection, but we do not have the same reason to reject the DBA for Conditionalization. In section 6 I discuss whether there are any other reasons to reject the DBA for Conditionalization, but first (in section 5) I contrast my new understanding DBAs with that of Rachael Briggs (Briggs 2009).

5. Briggs' 'Suppositional Test'

Briggs has also offered a new perspective on DBAs. As I do, she claims to be able to reject the DBA for Reflection without rejecting the DBA for Conditionalization. In this section I contrast my account with hers.

First I need to clear up a terminological difference between our accounts. I have aimed to give a new understanding of how DBAs work, and I have assumed that if a DBA *does* work (if an agent can be 'Dutch Booked') then the agent has been shown to be incoherent. In contrast, Briggs claims that some DBAs reveal incoherence, and some do not, and she applies a 'suppositional test' to differentiate between DBAs that reveal incoherence, and DBAs that don't. Briggs claims that an agent has been 'Dutch Booked' iff (s)he would accept as fair a set of bets that would result in a loss at every possible world where (s)he would accept those bets—i.e. at every possible world where his or her credence function (or some portion of it) is as it actually is. But (Briggs claims) showing that an agent can be Dutch Booked does not establish that (s)he is incoherent. This is where Briggs' suppositional test comes in: a DBA reveals incoherence iff the bets that the agent would accept as fair would lose him or her money at *every* possible world—including at worlds where the agent's credence function is different and so (s)he would accept different bets as fair.

> A set of bets reveals incoherence just in case at every possible world, the buyer of those bets loses more than he or she wins. But a set of bets counts as a Dutch book just in

> case at every possible world where the agent's beliefs condone the bets, the buyer of those bets loses more than he or she wins. So every set of bets that reveals incoherence counts as a Dutch book, but not every Dutch book reveals incoherence.
>
> (Briggs 2009: 80)

Briggs makes it clear elsewhere that 'possible world' here means 'suppositional world', and it is a consequence of Briggs' account that whatever evidence the agent gains is true at all suppositional worlds (Briggs 2009: 82). I have argued elsewhere that on this interpretation of 'possible world', Briggs' account draws the line between coherent and incoherent agents in the wrong place (Mahtani 2012), and I do not go through the argument for that here. We might instead try taking 'possible world' with its standard meaning, according to which a contingent claim (whether it is part of the agent's current evidence or not) holds at some but not all possible worlds. But then on this account the DBA for Conditionalization (as well as the DBA for Reflection) would fail. To see this, consider again our DBA against Fred. Suppose that in the actual world, E is true, so all 3 bets are offered. If we consider the outcome of these 3 bets at every possible world—including worlds at which E is false– we find that there are possible worlds where the 3 bets would give the agent an overall win. For example, take a possible world at which E is false, and H is true. At this world, bet J gives a loss of £0.04, bet K is called off, and bet L gives a profit of £0.10, resulting in an overall profit of £0.06 for the agent. Thus if we take Briggs' account, but take 'possible world' in its standard sense (rather than in the sense that Briggs intends), then we find that the DBA for Conditionalization does not go through—just as the DBA for Reflection does not go through—and so we are unable to draw a distinction between the DBAs for Conditionalization and Reflection.

Having clarified this point, I step back and compare my account with Briggs' more generally. The distinction between the accounts is parallel to a distinction between two different accounts of validity. On one account of validity, an argument is valid iff there is no possible world at which the premises are true and the conclusion false. Thus to assess validity, we take the premises and conclusion—with their actual meanings—to every possible world: iff there is no world at which the premises are true and the conclusion false, then the argument is valid. This corresponds to a Briggs-style take on DBAs.[6] We take the relevant book of bets that the agent would accept as fair—with the claims involved in the bets taking their actual meanings—to every possible world: iff there is no world where the agent avoids a loss, then the agent is incoherent. On another account of validity, an argument is valid iff there is no interpretation under which the premises are true and the conclusion false. Thus to assess validity, we hold the world fixed, and vary the interpretation: iff there is no interpretation under which the premises are true and the conclusion false, then the argument is valid. This corresponds to my take on DBAs. We take the relevant book of bets that the agent would accept as fair, and—holding the world fixed—we vary the interpretation of the claims involved in those bets: iff there is no interpretation under which the agent avoids a loss, then the agent is incoherent.

The different accounts of logical validity each have various advantages and disadvantages: for example, Volker Halbach gives us a good reason to focus on validity in terms of interpretations (Halbach 2010: 20), whereas John Etcthmendy gives some objections to an account of validity in such terms (Etchemendy 1999). Perhaps the parallel accounts of incoherence with respect to credence functions will inherit some of these advantages and disadvantages.[7] But at any rate the account of incoherence in terms of interpretations seems better suited than the account in terms of possible worlds for drawing a distinction between the principles of Conditionalization and Reflection.

6. Should We Accept the DBA for Conditionalization?

We have seen that on my understanding of DBAs, the DBA for Reflection fails—and for reasons that have nothing to do with the fact that it is a diachronic DBA. Thus we don't need to avoid all diachronic DBAs for fear of being stuck with the Reflection Principle. However there may be other reasons for rejecting the DBA for Conditionalization, and in this section I explain how my new understanding of DBAs leaves open the question of whether the DBA for Conditionalization goes through.

One reason for finding the principle of Conditionalization implausible is this: it seems to rule that coherent agents never forget anything. To see this, suppose that an agent has a credence function Cr at t_0, and $\mathrm{Cr}(\mathrm{M}) = 1$ (where M is the claim that the agent is eating meatballs for dinner at t_0). Let E be the conjunction of all the bits of evidence that the agent learns between t_0 and t_1—where t_1 is a year later than t_0. If none of this evidence bears on M, then plausibly $\mathrm{Cr}(\mathrm{M}/\mathrm{E}) = 1$. But the agent might well forget M over the course of the year, in which case his credence function at t_1, Cr_E, may assign a low value (say 0.1) to M. Thus $\mathrm{Cr}_\mathrm{E}(\mathrm{M}) \neq \mathrm{Cr}(\mathrm{M}/\mathrm{E})$, and the agent has failed to conditionalize. Nevertheless, this agent seems to be coherent.[8]

It might seem then that Conditionalization is not quite the right principle of diachronic coherence. Perhaps what is appealing about Conditionalization is that it requires that agents do not change their credences in a whimsical way, but rather in a predictable way in response to changes in evidence. But forgetting *is* a change in evidence, for it is a loss of evidence—and a coherent agent's credence function can reasonably change whenever he or she loses evidence. Perhaps then we should replace the principle of Conditionalization with an amended principle. To explain this amended principle, I first explore exactly what is meant by a claim of the form $\mathrm{Cr}_\psi(\phi) = v$.

We might take this claim to mean that if there is a time at which the only additional evidence the agent has gained (since his or her credence function was Cr) is ψ, then at this time the agent's credence in ϕ is v.[9] Note that a time at which the only additional evidence the agent has gained is ψ can be a time at which the agent has lost some evidence. Thus in the meatballs example above, the agent's credence at t_1 is Cr_E, because the only additional evidence she has gained is E—even though (s)he has lost some evidence as well. I now introduce Cr^ψ, which

is slightly different from Cr_{ψ}. The claim $\mathrm{Cr}^{\psi}(\phi) = v$ means that if there is a time at which the total evidence that the agent has is (ψ plus the total body of evidence the agent has at the time when his or her credence function is Cr), then at this time the agent's credence in ϕ is v. We can now give an amended version of the principle of Conditionalization: this principle will not require that $\mathrm{Cr}_{\psi}(\phi) = \mathrm{Cr}(\phi/\psi)$, but instead will require that $\mathrm{Cr}^{\psi}(\phi) = \mathrm{Cr}(\phi/\psi)$. This amended principle is not violated in the meatball example above, because at t_1 the agent lacks some evidence that (s)he had at t_0, and so the agent's credence function at t_1 is not Cr^{E}.

If this adapted principle of Conditionalization is preferred to the original principle, that may seem to create a problem for my view. After all, didn't the DBA for Conditionalization go through on my new understanding of DBAs? In fact, to get the DBA for Conditionalization to go through—on either the old or new understanding of DBAs—we needed to suppose that the bookie had a complex credence spreadsheet for the agent. This spreadsheet gave not just the values that the agent's current credence function (Cr) assigns to each claim ϕ that Cr assigns a value to, but also the value Cr_{ψ} assigns to ϕ for every pair of claims ϕ and ψ that Cr assigns values to. Now we can consider the question: should the bookie have access to this information?

In general, the information that we imagine the bookie having access to reflects our intuitions about what is relevant when assessing whether an agent is coherent. This was why, on the old understanding of DBAs, the bookie had information about the agent's credence function, but not about the rest of the world: whether the agent had a coherent credence function was thought to depend *just* on facts about the agent's credence function. And it is why, on my new understanding of DBAs, the bookie does not have information about which interpretation is correct. This reflects my view that whether a credence function is coherent does not depend on how the subject-specific terms in the claims involved are interpreted: we can work out whether a credence function is (synchronically) coherent just from the logical form of the claims, together with the values assigned to each.[10] So the question to consider here is this: is the value of $\mathrm{Cr}_{\psi}(\phi)$ (for any ϕ and ψ that Cr assigns a value to) relevant in assessing whether an agent is coherent? Or is it rather the value of $\mathrm{Cr}^{\psi}(\phi)$ that is relevant? To put the point informally: should the bookie be allowed to know what credence the agent will have in a claim ϕ should the agent acquire (just) evidence ψ? Or should the bookie instead only be allowed to know what credence the agent will have in a claim ϕ should the agent acquire (just) evidence ψ *and not lose any evidence*? Which sort of information—if either—is relevant to an assessment of the agent's coherence? The answer to this question will affect which diachronic coherence principles can be supported by a DBA.

Nothing I have said about my new way of understanding DBAs prejudges this issue. A benefit of the new way of understanding DBAs is that it decisively rules out the DBA for Reflection—thus removing a source of opposition to diachronic DBAs in general. This opens the way to a debate about whether the DBA for conditionalization—or some other diachronic principle of coherence—is successful.

Notes

[1] I would like to thank Cian Dorr, Jennifer Nagel, Lee Walters, Timothy Williamson, Alistair Wilson, all members of the Theoretical Work in Progress Group at Oxford, and an anonymous reviewer for Noûs for their invaluable feedback on various stages of this paper.

[2] Milne gives a slightly more complicated example, but his point here is essentially the same. With my simple example of Charlotte, it might be objected that the bookie would not know what the agent's credence is in L. The thought is that the bookie would get to know that $Cr(Cr(L) = 0.75) = 0.8$ (for how else could the bookie know to offer bet D?), but there is no reason for the bookie to know that $Cr(L) = 0.75$, as no bet is placed on L. But we can easily adapt the book of bets to avoid this objection, by adding two more bets to the book: these are bets on L that will 'cancel each other out', resulting jointly in neither a profit nor a loss for the agent.

[3] Thanks to Lee Walters for this way of putting the point.

[4] Whether the bookie should have information about the agent's credence function in *the past* (i.e. whether we should imagine the bookie remembering or forgetting the earlier values in the credence spreadsheet) is an interesting question—but it is not directly relevant to this assessment of the DBA for Reflection.

[5] For example, suppose that an agent is wondering whether (Q) she has a credence of exactly 0.5 in any claim. She is unsure whether or not Q is true—because (as can be the case even for coherent agents) she is not certain of all facts about her own credence function. She is also unsure of her own credence in Q, and she has some credence strictly between 0 and 1 that $Cr(Q) = 0.5$. It seems reasonable that $Cr(Q/Cr(Q) = 0.5) = 1$, in which case the agent violates the synchronic Reflection Principle, despite being coherent.

[6] As explained above, Briggs is using 'possible world' in a technical sense, so the analogy here is inexact.

[7] To see how my account might inherit some of the problems raised by Etchemendy, suppose that an agent has a credence of 0.6 that there are at least 2 things. This claim doesn't contain any subject-specific expressions: its logical form is: $\exists x \exists y\ x \neq y$. Thus the bookie can tell, from looking at the credence spreadsheet (logical form version) for the agent, that the agent has a credence of 0.6 in this claim. Given that the bookie is allowed to know other facts about how the world is, he can know that there *are* at least two things—and so he can offer the agent a bet that he knows the agent will lose (e.g. the agent gets £0.60, and pays out £1 iff there are at least 2 things). Thus on my new way of understanding DBAs, the agent has been shown to be incoherent. This is an unwelcome result: intuitively a coherent (but massively deluded agent) can be unsure whether there are at least 2 things.

To deal with this objection, I could drop the claim that the bookie can know facts about the world other than those contained in the credence spreadsheet (logical form version). Other than the information he has got from the spreadsheet, he knows nothing else that will allow him to rule out any logically possible world.

The parallel move for the account of validity in terms of interpretations faces the problem that the account is no longer reductive: the notion of 'logically possible' is left unexplained. But that is not a pressing problem for my account of coherence: I do not aim to give a reductive account of logical possibility.

[8] This example is adapted from Talbott 1991.

[9] This sentence is rather vague. If we try to make it precise various problems arise—which may indicate further problems with the principle of Conditionalization. If we take the sentence as a material conditional, then if there is *not* a time at which the only evidence the agent has gained is ψ, then the conditional is trivially true—no matter what number we substitute for v. Thus Cr_ψ will not be a function. And of course if the bookie gets to see whether Cr_ψ assigns lots of values to ϕ, or just one value, then he will at once know whether or not the agent learns that ψ—and so whether ψ is true. This will allow agents to be Dutch-Booked too easily. We might deal with this by ruling that if there is a time at which the only evidence the agent has gained is ψ, then $Cr_\psi(\phi)$ is the value that the agent assigns at this time to ϕ, and if there is not such a time, then $Cr_\psi(\phi)$ is 0—or some other randomly selected number. Alternatively, we might treat the conditional as a counterfactual. But all of these approaches may be at risk of giving the bookie too much information.

[10] Analogously, I claim that we can work out whether an agent's belief state is coherent just by looking at the logical forms of the set of claims believed.

References

Briggs, R. 2009. 'Distorted Reflection.' *Philosophical Review*, 118 (1), 59–85.
Christensen, D. 1991. 'Clever Bookies and Coherent Beliefs.' *Philosophical Review*, 100 (2), 229–247.
Etchemendy, J. 1990. *The Concept of Logical Consequence*, Harvard University Press.
Halbach, V. 2010. *The Logic Manual*, Oxford University Press.
Lewis, D. 1999. 'Why Conditionalize?' in Lewis, D. *Papers in Metaphysics and Epistemology*, CUP.
Mahtani, A. 2012. 'Diachronic Dutch Book Arguments.' *Philosophical Review*, 121 (3), 443–450.
Milne, P. 1991. 'A Dilemma for Subjective Bayesians—And How to Resolve It.' *Philosophical Studies*, 62 (3): 307–314.
Talbott, W. 1991. 'Two Principles of Bayesian Epistemology.' *Philosophical Studies*, 62, 135–50.
Van Fraassen, B. 1984. 'Belief and the Will.' *Journal of Philosophy*, 81 (5), 235–256.