<u>**Diachronic Dutch Books**</u>

Please note that this is an earlier version of the published paper – some changes were made after this.

Numerous counterexamples to Bas Van Fraassen's Reflection Principle (1984, 1995) show that an agent can violate this principle without incoherence.[1] However there is a Diachronic Dutch Book argument (DBA) to the conclusion that any agent who violates this principle is incoherent. Given that this argument gives the wrong result, should we distrust all Diachronic DBAs – including David Lewis' argument for Conditionalization (Lewis 1999)?

Rachael Briggs argues that Diachronic DBAs need not be rejected en masse, for she has a 'suppositional test' that marks out an important distinction between them. According to this suppositional test, agents who violate Conditionalization are incoherent whereas agents who violate Reflection may be merely self-doubting (Briggs 2009).

I argue that Briggs' suppositional test places the bar for coherence unreasonably high: certain cases of self-doubt are wrongly classified by the suppositional test as cases of incoherence. Thus – given that the suppositional test is unsatisfactory – we are left without any justification for discriminating between the diachronic DBAs.

Incoherence and Self-Doubt

The distinction between incoherence and self doubt can be brought out by contrasting an agent who believes 'P &not-P', with an agent who believes 'I believe (P & not-P)'. The content of the first agent's belief is false in every suppositional world, and so (according to Briggs) the first agent must be incoherent. In contrast the second agent may be merely self-doubting: he may be actually coherent whilst believing (falsely) that he is incoherent. His belief that he is incoherent (i.e. that he believes (P&not-P)) is itself a coherent belief, for its content is true at some suppositional worlds (Briggs 2009: 78-80).

Briggs makes a parallel distinction between an agent who fails to conditionalize (diachronically incoherent, according to Briggs), and an agent who violates the Reflection Principle. In typical counterexamples to the Reflection Principle, an agent suspects that she might in the future fail to conditionalize (e.g. she thinks that she might forget something, or that she might reason wrongly). Provided that such an agent does in fact conditionalize, Briggs classifies the agent as self-doubting

---

[1] Many of the counter-examples that Briggs describes (Briggs 2009: 64-66) need to be adjusted to be effective against the Reflection Principle. See footnote 2.

rather than diachronically incoherent. On Briggs' view, an agent can be coherent even if the following two conditions hold, where Cr_0 is her credence function at time t_0, and Cr_1 is her credence function at some later time t_1:

SUSPICION: $Cr_0(Cr_1(A)=r)=d$ where $0<d<1$

CONDITIONAL CREDENCE: $Cr_0(A/Cr_1(A)=r)=n$, where $r\neq n$

The Reflection Principle rules that an agent for whom both SUSPICION and CONDITIONAL CREDENCE hold is incoherent. Thus Briggs rejects the Reflection Principle, while endorsing Conditionalization. Correspondingly, she argues that we ought to be persuaded by the DBA for Conditionalization, but not by the DBA for Reflection.

The DBA for Reflection involves 3 bets, 2 of which are offered at t_0 and 1 of which is offered at t_1 iff $Cr_1(A)=r$. The bets are designed so that all the bets offered will be considered fair by an agent for whom SUSPICION and CONDITIONAL CREDENCE hold, and the payout of the bets is guaranteed to be $d(r-n)$. Thus given that $r\neq n$, the bets can be used to extract money from the agent (Briggs 2009: 64). Briggs points out, however, that the DBA has not shown that an agent who violates Reflection will accept a set of bets that will lose her money *in every suppositional world*. To see this, suppose that in the actual world $Cr_1(A) \neq r$ and the agent is offered and accepts only the first two bets; the payout of these bets at a suppositional world where $Cr_1(A)=r$ will not be $d(r-n)$. This is Briggs' 'suppositional test', and it rules that an agent who violates Reflection may be coherent.

The DBA for Conditionalization is superficially quite similar to the DBA for Reflection. It involves 3 bets, 2 of which are offered at t_0, and 1 of which is offered at t_1 iff the agent learns that E. At t_1 the agent will either learn E (and nothing else) or she will learn not-E. The bets are designed so that the payout is guaranteed to be $d(r-n)$, where $Cr_0(A/E)=n$, $Cr_E(A)=r$ and $Cr_0(E)=d$: thus these bets can be used to extract money from any agent who violates conditionalization – i.e. for whom $r\neq n$ (Briggs 2009: 63). We might think that we could now make a move parallel to that made above in the discussion of the DBA for Reflection. For example we could suppose that in the actual case the agent does not learn E, and so the 3rd bet is not made. We could then point to a suppositional world at which E is true, where the payout of the two bets actually made will not be $d(r-n)$. According to Briggs, however, there will be no such suppositional world: 'I'll adopt Lewis's assumption that the agent stands no chance of mistaking her evidence, so that if the agent learns E [or not-E], then all the suppositional worlds must be ones where E is true [or not true]' (Briggs 2009: 82).

A High Bar for Coherence

Briggs is using 'suppositional' in such a way that whatever counts as evidence is true at all suppositional worlds. Clearly then, the worlds that are suppositional for an agent at a time need not be suppositional for another agent, or even for the same agent at a different time. In Briggs' discussion of the DBA for Conditionalization, she assumes that what is evidence for the agent at the actual world at t_1 must be true at all suppositional worlds in the relevant sense, so I will also adopt this assumption.

Let us turn then to Briggs' discussion of the DBA for the Reflection Principle. Briggs claims that if at the actual world $Cr_1(A) \neq r$, then there will nevertheless be suppositional worlds at which $Cr_1(A)=r$. This assumes that $Cr_1(A) \neq r$ is not evidence for the agent at the actual world at t_1. Of course it is quite possible for an agent to have a credence other than r in A at t_1 and not know that she does, for we do not always know what our current credences are. Equally however we sometimes do have some knowledge about our own current credences: an agent could have a credence other than r in A at t_1, and know that she did. In a case where an agent does have this sort of self-knowledge, presumably this would count as evidence for her, and so would be true across all suppositional worlds.

Consider then this Adapted Reflection Principle: an agent is incoherent if, in addition to SUSPICION and CONDITIONAL CREDENCE, SELF-KNOWLEDGE also holds:

SELF-KNOWLEDGE $(Cr_1(A)=r \rightarrow Cr_1(Cr_1(A)=r)=1)$ & $(Cr_1(A) \neq r \rightarrow Cr_1(Cr_1(A) \neq r)=1)$

SELF-KNOWLEDGE states that the agent will know, at t_1, whether or not her credence in A is r: this will be part of her evidence at t_1. The arrows in this claim should be understood as material conditionals. The DBA for the original Reflection Principle can be used to argue for the Adapted Reflection Principle. I show that an agent who violates the Adapted Reflection Principle will accept as fair a set of bets that is guaranteed to lose her money. Furthermore, the set of bets that she accepts as fair in the actual world will lose her money in every suppositional world. Thus – by Briggs' suppositional test – such an agent must be incoherent.

To run the DBA for the Adapted Reflection Principle, we take an agent for whom SELF-KNOWLEDGE, SUSPICION and CONDITIONAL CREDENCE all hold. We then offer the agent the same collection of bets that were offered to the agent in the original DBA for Reflection – i.e. bets 1 and 2 at t_0, and bet 3 at t_1 iff $Cr_1(A)=r$. I reproduce these bets here:

		Payout			*Payout*
Bet 1	A & $Cr_1(A)=r$	$1-n$	Bet 2	$Cr_1(A)=r$	$(d-1)(r-n)$
	~A & $Cr_1(A)=r$	$-n$		$Cr_1(A) \neq r$	$d(r-n)$
	$Cr_1(A) \neq r$	0			

		Payout
Bet 3	A	$r-1$
	~A	r

If $Cr_1(A)=r$, and the agent accepts all 3 bets, then (by SELF-KNOWLEDGE) the agent will know $Cr_1(A)=r$ at t_1 and so this will hold in every suppositional world; similarly, if $Cr_1(A)\neq r$, and the agent accepts only the bets 1 and 2, then (by SELF-KNOWLEDGE) the agent will know $Cr_1(A)\neq r$ at t_1, and so this will hold at every suppositional world. Thus the bets that the agent accepts in the actual world are guaranteed to payout $d(r-n)$ in every suppositional world, and so – by Briggs' suppositional test – an agent who violates the Adapted Reflection Principle is incoherent.

Briggs' suppositional test leads us astray here, for an agent who violates the Adapted Reflection Principle need not be incoherent. To see this, recall that (as Briggs rightly stresses) an agent can – coherently – suspect at t_0 that she might end up by some process other than conditionalization with a credence of r in A at t_1. Further, an agent can have at t_0 a credence other than r in A conditional on this possibility: in other words, an agent can coherently violate the Reflection Principle. An agent who violates the Adapted Reflection Principle must satisfy the following further condition: she must know at t_1 whether or not her credence in A at t_1 is r. It is hard to see how satisfying this additional condition could give rise to incoherence.

I illustrate this point by adapting one of the counterexamples to the Reflection Principle that Briggs describes. Take this (adjusted)[2] case from William Talbott (summarised in Briggs 2009: 65): Let S be the claim that you are, at t_0, eating a dinner of spaghetti and meatballs. At t_0 you suspect that you might have forgotten S by t_1, in which case you will remember that you eat spaghetti for dinner 10 percent of the time, and $Cr_1(S)$ will be 0.10. Your credence at t_0 in S conditional on $Cr_1(S)=0.10$ is much higher than 0.10. Thus SUSPICION and CONDITIONAL CREDENCE hold, and so you violate the

---

[2] In the original case, the agent was certain that she would forget S by t_1 – that is, she was certain that she would fail to conditionalize – and so Briggs ought to count the agent in the original case as incoherent. Thus the mere fact that the Reflection Principle would also rule that this agent is incoherent is no reason for Briggs to reject the Reflection Principle.

The case only works as a counter-example to a certain narrow-scope reading of the Reflection Principle, according to which if the agent knows at t_0 that $Cr_1(S)=0.10$, then at t_0 his credence in S *ought* to be 0.10. As Briggs says, this is the 'wrong advice' (Briggs 2009: 65), and so the example shows that the narrow-scope reading ought to be rejected. In contrast, a (more plausible) wide-scope reading would rule that the agent ought not to be such that he both knows at t_0 that $Cr_1(S)=0.10$ and has a credence at t_0 in S that is other than 0.10. This is something that Briggs would agree with, for she too would claim that the agent is incoherent, and so the case does not function for her as a counterexample to the wide-scope Reflection Principle.

In my adjusted case, the agent is not certain that she will fail to conditionalize. Thus Briggs would claim that she may be coherent, whereas Van Fraassen's Reflection Principle would class her as incoherent.

Reflection Principle whilst (on Briggs view) remaining coherent. We can convert this case into a counterexample to the Adapted Reflection Principle by adding simply that at t_1 you will know whether $Cr_1(S)=0.10$, and so guaranteeing that the relevant SELF-KNOWLEDGE statement holds. This is a plausible addition: we can suppose that at t_1 you will be offered the chance to bet on S, and your response will reveal your own credence. If (as Briggs maintains) you are a coherent agent in the original version of the scenario, then merely adding the relevant SELF-KNOWLEDGE claim does not make you incoherent.

We could adapt other counterexamples to the original Reflection Principle in a similar way. Thus the Adapted Reflection Principle is no more acceptable than the Reflection Principle. An agent who violates the Adapted Reflection Principle ought to be classed as self-doubting, rather than incoherent. Yet Briggs' suppositional test places this agent on the wrong side of the divide.

Being a Conditionalizer

Briggs has stated in personal correspondence that she is willing to bite the bullet on this issue and classify violations of the Adapted Reflection Principle as cases of incoherence, along with violations of Conditionalization. The thought is that an agent who violates the Adapted Reflection Principle must either be synchronically incoherent at t_0, or must not be a 'conditionalizer', where a conditionalizer is a person who knows for certain that she will conditionalize. As Briggs clarifies: 'You're not a conditionalizer as long as you *might*, for all you know, fail to conditionalize. It's not good enough to end up having conditionalized by sheer luck' (Personal Correspondence). It is indeed true that if you are a conditionalizer in this sense, and if various other background assumptions hold, then you will obey the Adapted Reflection Principle. Briggs proves a related claim in her argument for 'Qualified Reflection' (Briggs 2009: 69). However, to demand that every coherent agent must be a 'conditionalizer' in this sense is to set the bar for coherence too high.

To see this, consider an agent watching a game of chess. At a certain point in the game, she is told by an extremely reliable chess master that one player can definitely check-mate the other in two moves. The information that our agent has about the rules of the game would allow her to calculate which player this is – but she cannot be bothered to work it out. Instead she guesses, and luckily she guesses right. Such an agent might be classed as irrational, but it is far from obvious that such an agent should be classed as *incoherent*.[3] We can make this point more compelling by noting that there is a wide gulf between being certain that you will conditionalize and simply guessing. Our

---

[3] Thanks to an anonymous reviewer for suggesting this point.

agent might work through the possible moves quickly to arrive at her (correct) answer. She might feel confident that she has reasoned correctly without being quite certain: for all she knows, she *might* have failed to conditionalise. This agent is thus not a conditionalizer in Briggs sense, but it would not be reasonable to classify her as incoherent.

In any case, to classify as incoherent an agent who merely suspects that she might fail to conditionalize seems counter to Briggs' wider aim: to mark a distinction between an agent who fails to conditionalize (incoherent), and an agent who violates Reflection (self-doubting). As Briggs states: "According to the suppositional test ... violating conditionalization is a type of incoherence, while violating Reflection is a type of self-doubt. This result makes sense.... to violate Reflection is to suspect one will fail to conditionalize – that is, to suspect oneself of diachronic incoherence" (Briggs 2009: 82).

 The suppositional test rules that an agent who violates the Adapted Reflection Principle is incoherent. This result does *not* make sense, for an agent who violates the Adapted Reflection Principle – like the agent who violates the original Reflection Principle – need not actually fail to conditionalize, but may merely suspect herself of diachronic incoherence.

Conclusion

Briggs proposed a suppositional test that distinguished the DBA for Conditionalization from the DBA for Reflection. Unfortunately, the test sets the bar for coherence too high, wrongly classifying agents who violate my Adapted Reflection Principle as incoherent. Briggs' suppositional test must therefore be rejected, and it seems that the diachronic DBAs for Conditionalization and Reflection must stand – or rather fall – together.

Bibliography

Briggs, Rachael. 2009. 'Distorted Reflection.' *Philosophical Review* 118: 59-85

Lewis, David. 1999. 'Why Conditionalize?' *Papers in Metaphysics and Epistemology*, 403-7. Cambridge: Cambridge University Press

Van Fraassen, Bas. 1984. 'Belief and the Will.' *Journal of Philosophy* 81: 235-56.

----. 1995. 'Belief and the Problem of the Ulysses and the Sirens.' *Philosophical Studies* 77: 7-37.