

# Games, Beliefs and Credences

Brian Weatherson

---

In previous work (Weatherson, 2005, 2011, 2012b) I've defended an interest-relative theory of belief. This paper continues the defence. I have four aims.

1. To offer a new kind of reason for being unsatisfied with the simple Lockean reduction of belief to credence.
2. To defend the legitimacy of appealing to credences in a theory of belief.
3. To illustrate the importance of theoretical, as well as practical, interests in an interest-relative account of belief.
4. To have another try at extending my basic account of belief to cover propositions that are practically and theoretically irrelevant to the agent.

You're probably familiar with the following dialectic. We want there to be some systematic connection between credences and beliefs. At first blush, saying that a person believes  $p$  and has a very low credence in  $p$  isn't just an accusation of irrationality, it is literally incoherent. The simplest such connection would be a reduction of beliefs to credences. But the simplest reductions don't work.

If we identify beliefs with credence 1, and take credences to support betting dispositions, then a rational agent will have very few beliefs. There are lots of things that an agent, we would normally say, believes even though she wouldn't bet on them at absurd odds. Note that this argument doesn't rely on *reducing* credences to betting dispositions; as long as credences support the betting dispositions, the argument goes through.

A simple retreat is to the so-called **Lockean thesis**, which holds that to believe that  $p$  is to have credence in  $p$  greater than some threshold  $t$ , where  $t < 1$ . Just how the threshold is determined could be a matter of some discretion. Perhaps it is a function of the agent's situation, or of the person ascribing beliefs to the agent, or to the person evaluating that ascription. Never mind these complexities; assuming all such things are held fixed, the Lockean thesis says that there is a threshold  $t$  such that everything with credence above  $t$  is believed.

---

<sup>†</sup> In progress. Draft intended for the 2013 Formal Epistemology Workshop/Rutgers Epistemology Conference. Note for those attending those conferences: Obviously this is too long for me to read all of it at the conference; I'll be doing edited highlights. And some of the material, especially the stuff about updating at the end of section 1, and about encoding toward the end of section 3, could use with being more crisply stated. Hopefully I'll have made improvements on those fronts by May.

There's a simple objection to the Lockean thesis. Given some very weak assumptions about the world, it implies that there are plenty of quadruples  $\langle S, A, B, A \wedge B \rangle$  such that

- $S$  is a rational agent.
- $A, B$  and  $A \wedge B$  are propositions.
- $S$  believes  $A$  and believes  $B$ .
- $S$  does not believe  $A \wedge B$ .
- $S$  knows that she has all these states, and consciously reflectively endorses them.

Now one might think, indeed I do think, that such quadruples do not exist at all. But set that objection aside. If the Lockean is correct, these quadruples should be everywhere. That's because for any  $t \in (0, 1)$  you care to pick, quadruples of the form  $\langle S, C, D, C \wedge D \rangle$  are very very common.

- $S$  is a rational agent.
- $C, D$  and  $C \wedge D$  are propositions.
- $S$ 's credence in  $C$  is greater than  $t$ , and her credence in  $D$  is greater than  $t$ .
- $S$ 's credence in  $C \wedge D$  is less than  $t$ .
- $S$  knows that she has all these states, and reflectively endorses them.

The best arguments for the existence of quadruples  $\langle S, A, B, A \wedge B \rangle$  are non-constructive existence proofs. David Christensen (2005) for instance, argues from the existence of the preface paradox to the existence of these quadruples. I've expressed some reservations about that argument in the past (Weatherson, 2005). But what I want to stress here is that even if these existence proofs work, they don't really prove what the Lockean needs. They don't show that quadruples satisfying the constraints we associated with  $\langle S, A, B, A \wedge B \rangle$  are just as common as quadruples satisfying the constraints we associated with  $\langle S, C, D, C \wedge D \rangle$ , for any  $t$ . But if the Lockean were correct, they should be exactly as common.

This kind of consideration pushes some of us, well me in any case, towards an interest-relative account of belief. But I'm going to set that worry aside to start by investigating a different objection. This objection holds that the Lockean thesis could not be true, because credence 1 is not *sufficient* for belief. That is, the Lockean is committed to the thesis known as *regularity*; that everything left open by belief gets a positive credence. I think regularity is false. That's hardly news, there are plenty of good arguments against it, though most of these

involve cases with some idealisations. Timothy Williamson (2007a) has a compelling argument against regularity turning on reflections about a case involving infinite coin flips.<sup>1</sup> I'm going to offer a 'finite' argument against regularity, which I hope is of independent interest, and from that conclude the Lockean is mistaken. There is a worry that my argument against the Lockean also undermines my preferred positive view, and I'll suggest an independently motivated patch. I'll then turn to Richard Holton's attack on the very notion of credence, which obviously would have repercussions for attempts to understand beliefs in terms of credences were it to succeed. I think it doesn't succeed, but it does show there are important and underappreciated constraints on a theory of belief. I'll conclude with a comparison between my preferred interest-relative account of belief, and a recent account suggested by Jacob Ross and Mark Schroeder. The short version of the comparison is that I think there's less difference between the views than Ross and Schroeder think, though naturally I think what differences there are favour my view.

## 1 Playing Games with a Lockean

I'm going to raise problems for Lockeans, and for defenders of regularity in general, by discussing a simple game. The game itself is a nice illustration of how a number of distinct solution concepts in game theory come apart. (Indeed, the use I'll make of it isn't a million miles from the use that Kohlberg and Mertens (1986) make of it.) To set the problem up, I need to say a few words about how I think of game theory. This won't be at all original - most of what I say is taken from important works by Robert Stalnaker (1994, 1996, 1998, 1999). But it is different to what I used to think, and perhaps to what some other people think too, so I'll set it out slowly.<sup>2</sup>

Start with a simple decision problem, where the agent has a choice between two acts  $A_1$  and  $A_2$ , and there are two possible states of the world,  $S_1$  and  $S_2$ , and the agent knows the payouts for each act-state pair are given by the following table.

---

<sup>1</sup>If there's any gap in Williamson's argument, it is I think at the point where he concludes that any two infinite sequences of coin flips have the same probability of landing all heads. I think that the defender of non-numerical, comparative approaches to probability can deny that with some plausibility. Perhaps the two sequences of coin flips have *incomparable* probabilities of landing all heads. But this leads us into complications that are irrelevant to this paper, especially since I think it turns out there is a sound Williamsonian argument against the Lockean who lets different sequences have incomparable probabilities.

<sup>2</sup>I'm grateful to the participants in a game theory seminar at Arché in 2011, especially Josh Dever and Levi Spectre, for very helpful discussions that helped me see through my previous confusions.

|       |       |       |
|-------|-------|-------|
|       | $S_1$ | $S_2$ |
| $A_1$ | 4     | 0     |
| $A_2$ | 1     | 1     |

What to do? I hope you share the intuition that it is radically underdetermined by the information I've given you so far. If  $S_2$  is much more probable than  $S_1$ , then  $A_2$  should be chosen; otherwise  $A_1$  should be chosen. But I haven't said anything about the relative probability of those two states. Now compare that to a simple game. Row has two choices, which I'll call  $A_1$  and  $A_2$ . Column also has two choices, which I'll call  $S_1$  and  $S_2$ . It is common knowledge that each player is rational, and that the payouts for the pairs of choices are given in the following table. (As always, Row's payouts are given first.)

|       |       |       |
|-------|-------|-------|
|       | $S_1$ | $S_2$ |
| $A_1$ | 4, 0  | 0, 1  |
| $A_2$ | 1, 0  | 1, 1  |

What should Row do? This one is easy. Column gets 1 for sure if she plays  $S_2$ , and 0 for sure if she plays  $S_1$ . So she'll play  $S_2$ . And given that she's playing  $S_2$ , it is best for Row to play  $A_2$ .

You probably noticed that the game is just a version of the decision problem that we discussed a couple of paragraphs ago. The relevant states of the world are choices of Column. But that's fine; we didn't say in setting out the decision problem what constituted the states  $S_1$  and  $S_2$ . And note that we solved the problem without explicitly saying anything about probabilities. What we added was some information about Column's payouts, and the fact that Column is rational. From there we deduced something about Column's play, namely that she would play  $S_2$ . And from that we concluded what Row should do.

There's something quite general about this example. What's distinctive about game theory isn't that it involves any special kinds of decision making. Once we get the probabilities of each move by the other player, what's left is (mostly) expected utility maximisation. (We'll come back to whether the 'mostly' qualification is needed below.) The distinctive thing about game theory is that the probabilities aren't specified in the setup of the game; rather, they are solved for. Apart from special cases, such as where one option strictly dominates another, we can't say much about a decision problem with unspecified probabilities. But we can and do say a lot about games where the setup of the game doesn't specify the probabilities, because we can solve for them given the other information we have.

This way of thinking about games makes the description of game theory as ‘interactive epistemology’ (Aumann, 1999) rather apt. The theorist’s work is to solve for what a rational agent should think other rational agents in the game should do. From this perspective, it isn’t surprising that game theory will make heavy use of equilibrium concepts. In solving a game, we must deploy a theory of rationality, and attribute that theory to rational actors in the game itself. In effect, we are treating rationality as something of an unknown, but one that occurs in every equation we have to work with. Not surprisingly, there are going to be multiple solutions to the puzzles we face.

This way of thinking lends itself to an epistemological interpretation of one of the most puzzling concepts in game theory, the mixed strategy. Arguably the core solution concept in game theory is the Nash equilibrium. As you probably know, a set of moves is a Nash equilibrium if no player can improve their outcome by deviating from the equilibrium, conditional on no other player deviating. In many simple games, the only Nash equilibria involve mixed strategies. Here’s one simple example.

|       |       |       |
|-------|-------|-------|
|       | $S_1$ | $S_2$ |
| $A_1$ | 0, 1  | 10, 0 |
| $A_2$ | 9, 0  | -1, 1 |

This game is reminiscent of some puzzles that have been much discussed in the decision theory literature, namely asymmetric Death in Damascus puzzles. Here Column wants herself and Row to make the ‘same’ choice, i.e.,  $A_1$  and  $S_1$  or  $A_2$  and  $S_2$ . She gets 1 if they do, 0 otherwise. And Row wants them to make different choices, and gets 10 if they do. Row also dislikes playing  $A_2$ , and this costs her 1 whatever else happens. It isn’t too hard to prove that the only Nash equilibrium for this game is that Row plays a mixed strategy playing both  $A_1$  and  $A_2$  with probability  $1/2$ , while Column plays the mixed strategy that gives  $S_1$  probability  $11/20$ , and  $S_2$  with probability  $9/20$ .

Now what is a mixed strategy? It is easy enough to take away from the standard game theory textbooks a **metaphysical** interpretation of what a mixed strategy is. Here, for instance, is the paragraph introducing mixed strategies in Dixit and Skeath’s *Games of Strategy*.

When players choose to act unsystematically, they pick from among their pure strategies in some random way . . . We call a random mixture between these two pure strategies a mixed strategy. (Dixit and Skeath, 2004, 186)

Dixit and Skeath are saying that it is definitive of a mixed strategy that players use some kind of randomisation device to pick their plays on any particular run of a game. That is, the probabilities in a mixed strategy must be in the world; they must go into the players' choice of play. That's one way, the paradigm way really, that we can think of mixed strategies metaphysically.

But the understanding of game theory as interactive epistemology naturally suggests an **epistemological** interpretation of mixed strategies.

One could easily ... [model players] ... turning the choice over to a randomizing device, but while it might be harmless to permit this, players satisfying the cognitive idealizations that game theory and decision theory make could have no motive for playing a mixed strategy. So how are we to understand Nash equilibrium in model theoretic terms as a solution concept? We should follow the suggestion of Bayesian game theorists, interpreting mixed strategy profiles as representations, not of players' choices, but of their beliefs. (Stalnaker, 1994, 57-8)

One nice advantage of the epistemological interpretation, as noted by Binmore (2007, 185) is that we don't require players to have  $n$ -sided dice in their satchels, for every  $n$ , every time they play a game.<sup>3</sup> But another advantage is that it lets us make sense of the difference between playing a pure strategy and playing a mixed strategy where one of the 'parts' of the mixture is played with probability one.

With that in mind, consider the below game, which I'll call RED-GREEN. I've said something different about this game in earlier work (Weatherson, 2012a). But I now think that to understand what's going on, we need to think about mixed strategies where one element of the mixture has probability one.

Informally, in this game  $A$  and  $B$  must each play either a green or red card. I will capitalise  $A$ 's moves, i.e.,  $A$  can play GREEN or RED, and italicise  $B$ 's moves, i.e.,  $B$  can play *green* or *red*. If two green cards, or one green card and one red card are played, each player gets \$1. If two red cards are played, each gets nothing. Each cares just about their own wealth, so getting \$1 is worth 1 util. All of this is common knowledge. More formally, here is the game table, with  $A$  on the row and  $B$  on the column.

---

<sup>3</sup>Actually, I guess it is worse than if some games have the only equilibria involving mixed strategies with irrational probabilities. And it might be noted that Binmore's introduction of mixed strategies, on page 44 of his (2007), sounds much more like the metaphysical interpretation. But I think the later discussion is meant to indicate that this is just a heuristic introduction; the epistemological interpretation is the correct one.

|       | <i>green</i> | <i>red</i> |
|-------|--------------|------------|
| GREEN | 1, 1         | 1, 1       |
| RED   | 1, 1         | 0, 0       |

When I write game tables like this, and I think this is the usual way game tables are to be interpreted (Weatherson, 2012b), I mean that the players know that these are the payouts, that the players know the other players to be rational, and these pieces of knowledge are common knowledge to at least as many iterations as needed to solve the game. With that in mind, let's think about how the agents should approach this game.

I'm going to make one big simplifying assumption at first. We'll relax this later, but it will help the discussion a lot I think to start with this assumption. This assumption is that the doctrine of **Uniqueness** applies here; there is precisely one rational credence to have in any salient proposition about how the game will play. Some philosophers think that Uniqueness always holds (White, 2005). I don't, but it does seem like it might *often* hold. Anyway, I'm going to start by assuming that it does hold here.

The first thing to note about the game is that it is symmetric. So the probability of *A* playing GREEN should be the same as the probability of *B* playing *green*, since *A* and *B* face exactly the same problem. Call this common probability  $x$ . If  $x < 1$ , we get a quick contradiction. The expected value, to Row, of GREEN, is 1. Indeed, the known value of GREEN is 1. If the probability of *green* is  $x$ , then the expected value of RED is  $x$ . So if  $x < 1$ , and Row is rational, she'll definitely play GREEN. But that's inconsistent with the claim that  $x < 1$ , since that means that it isn't definite that Row will play GREEN.

So we can conclude that  $x = 1$ . Does that mean we can know that Row will play GREEN? No. Assume we could conclude that. Whatever reason we would have for concluding that would be a reason for any rational person to conclude that Column will play *green*. Since any rational person can conclude this, Row can conclude it. So Row knows that she'll get 1 whether she plays GREEN or RED. But then she should be indifferent between playing GREEN and RED. And if we know she's indifferent between playing GREEN and RED, and our only evidence for what she'll play is that she's a rational player who'll maximise her returns, then we can't be in a position to know she'll play GREEN.

I think the arguments of the last two paragraphs are sound. We'll turn to an objection presently, but let's note how bizarre is the conclusion we've reached. One argument has shown that it could not be more probable that Row will play GREEN. A second argument has shown that we can't know that Row will play GREEN. It reminds me of examples involving blindspots (Sorensen, 1988). Consider this case:

(B) Brian does not know (B).

That's true, right? Assume it's false, so I do know (B). Knowledge is factive, so (B) is true. But that contradicts the assumption that it's false. So it's true. But I obviously don't know that it's true; that's what this very true proposition says.

Now I'm not going to rest anything on this case, because there are so many tricky things one can say about blindspots, and about the paradoxes generally. It does suggest that there are other finite cases where one can properly have maximal credence in a true proposition without knowledge.<sup>4</sup> And, assuming that we shouldn't believe things we know we don't know, that means we can have maximal credence in things we don't believe. All I want to point out is that this phenomena of maximal credence without knowledge, and presumably without full belief, isn't a quirky feature of self-reference, or of games, or of puzzles about infinity; it comes up in a wide range of cases.

For the rest of this section I want to reply to one objection, and weaken an assumption I made earlier. The objection is that I'm wrong to assume that agents will only maximise expected utility. They may have tie-breaker rules, and those rules might undermine the arguments I gave above. The assumption is that there's a uniquely rational credence to have in any given situation.

I argued that if we knew that *A* would play GREEN, we could show that *A* had no reason to play GREEN. But actually what we showed was that the expected utility of playing GREEN would be the same as playing RED. Perhaps *A* has a reason to play GREEN, namely that GREEN weakly dominates RED. After all, there's one possibility on the table where GREEN does better than RED, and none where RED does better. And perhaps that's a reason, even if it isn't a reason that expected utility considerations are sensitive to.

Now I don't want to insist on expected utility maximisation as the only rule for rational decision making. Sometimes, I think some kind of tie-breaker procedure is part of rationality. In the papers by Stalnaker I mentioned above, he often appeals to this kind of weak dominance reasoning to resolve various hard cases. But I don't think weak dominance provides a reason to play GREEN in this particular case. When Stalnaker says that agents should use weak dominance reasoning, it is always in the context of games where the agents' attitude towards the

---

<sup>4</sup>As an aside, the existence of these cases is why I get so irritated when epistemologists try to theorise about 'Gettier Cases' as a class. What does (B) have in common with inferences from a justified false belief, or with otherwise sound reasoning that is ever so close to issuing in a false conclusion due to relatively bad luck? As far as I can tell, the class of justified true beliefs that aren't knowledge is a disjunctive mess, and this should matter for thinking about the nature of knowledge. For further examples, see Williamson (2012).



game matrix is different to their attitude towards each other. One case that Stalnaker discusses in detail is where the game table is common knowledge, but there is merely common (justified, true) belief in common rationality. Given such a difference in attitudes, it does seem there's a good sense in which the most salient departure from equilibrium will be one in which the players end up somewhere else on the table. And given that, weak dominance reasoning seems appropriate.

But that's not what we've got here. Assuming that rationality requires playing GREEN/*green*, the players know we'll end up in the top left corner of the table. There's no chance that we'll end up elsewhere. Or, perhaps better, there is just as much chance we'll end up 'off the table', as that we'll end up in a non-equilibrium point on the table. To make this more vivid, consider the 'possibility' that *B* will play *blue*, and if *B* plays *blue*, *A* will receive 2 if she plays RED, and -1 if she plays GREEN. Well hold on, you might think, didn't I say that *green* and *red* were the only options, and this was common knowledge? Well, yes, I did, but if the exercise is to consider what would happen if something the agent knows to be true doesn't obtain, then the possibility that one agent will play blue certainly seems like one worth considering. It is, after all, a metaphysical possibility. And if we take it seriously, then it isn't true that under *any* possible play of the game, GREEN does better than RED.

We can put this as a dilemma. Assume, for *reductio*, that GREEN/*green* is the only rational play. Then if we restrict our attention to possibilities that are epistemically open to *A*, then GREEN does just as well as RED; they both get 1 in every possibility. If we allow possibilities that are epistemically closed to *A*, then the possibility where *B* plays *blue* is just as relevant as the possibility that *B* is irrational. After all, we stipulated that this is a case where rationality is common knowledge. In neither case does the weak dominance reasoning get any purchase.

With that in mind, we can see why we don't need the assumption of Uniqueness. Let's play through how a failure of Uniqueness could undermine the argument. Assume, again for *reductio*, that we have credence  $\varepsilon > 0$  that *A* will play RED. Since *A* maximises expected utility, that means *A* must have credence 1 that *B* will play *green*. But this is already odd. Even if you think people can have different reactions to the same evidence, it is odd to think that one rational agent could regard a possibility as infinitely less likely than another, given isomorphic evidence. And that's not all of the problems. Even if *A* has credence 1 that *B* will play *green*, it isn't obvious that playing RED is rational. After all, relative to the space of epistemic possibilities, GREEN weakly dominates RED. Remember that we're no longer assuming that it can be known what *A* or *B* will play. So even without Uniqueness, there are two reasons to think that it is wrong to have credence  $\varepsilon > 0$  that *A* will play RED. So we've still shown that credence

1 doesn't imply knowledge, and since the proof is known to us, and full belief is incompatible with knowing that you can't know, this is a case where credence 1 doesn't imply full belief. So whether *A* plays GREEN, like whether the coin will ever land tails, is a case the Lockean cannot get right, no matter where they set the threshold for belief; our credence is above the threshold, but we don't believe.

So I think this case is a real problem for a Lockean view about the relationship between credence and belief. If *A* is rational, she can have credence 1 that *B* will play *green*, but won't believe that *B* will play *green*. But now you might worry that my own account of the relationship between belief and credence is in just as much trouble. After all, I said that to believe *p* is, roughly, to have the same attitudes towards all salient questions as you have conditional on *p*. And it's hard to identify a question that rational *A* would answer differently upon conditionalising on the proposition that *B* plays *green*.

I think what went wrong in my earlier view was that I'd too quickly equated updating with conditionalisation. The two can come apart. Here's an example from Gillies (2010) that makes the point well.

I have lost my marbles. I know that just one of them – Red or Yellow – is in the box. But I don't know which. I find myself saying things like ... "If Yellow isn't in the box, the Red must be." (4:13)

As Gillies goes on to point out, this isn't really a problem for the Ramsey test view of conditionals.

The Ramsey test – the schoolyard version, anyway – is a test for when an indicative conditional is acceptable given your beliefs. It says that (if *p*)(*q*) is acceptable in belief state *B* iff *q* is acceptable in the derived or subordinate state *B*-plus-the-information-that-*p*. (4:27)

And he notes that this can explain what goes on with the marbles conditional. Add the information that Yellow isn't in the box, and it isn't just true, but must be true, that Red is in the box.

Note though that while we can explain this conditional using the Ramsey test, we can't explain it using any version of the idea that probabilities of conditionals are conditional probabilities. The probability that Red must be in the box is 0. The probability that Yellow isn't in the box is not 0. So conditional on Yellow not being in the box, the probability that Red must be in the box is still 0. Yet the conditional is perfectly assertable.

There is, and this is Gillies's key point, something about the behaviour of modals in the consequents of conditionals that we can't capture using conditional

probabilities, or indeed many other standard tools. And what goes for consequents of conditionals goes for updated beliefs too. Learn that Yellow isn't in the box, and you'll conclude that Red must be. But that learning can't go via conditionalisation; just conditionalise on the new information and the probability that Red must be in the box goes from 0 to 0.

Now it's a hard problem to say exactly how this alternative to updating by conditionalisation should work. But very roughly, the idea is that at least some of the time, we update by eliminating worlds from the space of possibilities. This affects dramatically the probability of propositions whose truth is sensitive to which worlds are in the space of possibilities.

For example, in the game I've been discussing, we should believe that rational *B* might play *red*. Indeed, the probability of that is, I think, 1. And whether or not *B* might play red is highly salient; it matters to the probability of whether *A* will play GREEN or RED. Conditionalising on something that has probability 1, such as that *B* will play *green*, can hardly change that probability. But updating on the proposition that *B* will play *green* can make a difference. We can see that by simply noting that the conditional *If B plays green, she might play red* is incoherent.

So I conclude that a theory of belief like mine can handle the puzzle this game poses, as long as it distinguishes between conditionalising and updating, in just the way Gillies suggests. To believe that *p* is to be disposed to not change any attitude towards a salient question on updating that *p*. (Plus some bells and whistles to deal with propositions that are not relevant to salient questions. We'll return to them below.) Updating often goes by conditionalisation, so we can often say that belief means having attitudes that match unconditionally and conditionally on *p*. But not all updating works that way, and the theory of belief needs to acknowledge this.

## 2 Holton on Credence

While I don't agree with the Lockeans, I do endorse a lot of similar theses to them about the relationship between belief and credence. These theses include that both beliefs and credences exist and that the two are constitutively (as opposed to merely causally) connected. I differ from the Lockeans in holding that both belief and credence have important explanatory roles, and that the connection between the two goes via the interests of the agent. As with most work in this area, my views start off from considerations of cases much like DeRose's famous bank cases.<sup>5</sup> Here's another contribution to the genre. I know it's an

---

<sup>5</sup>The idea of using allergies to illustrate the kind of case we're interested in is due to Ross and Schroeder (fort), and I'm grateful for the idea. It makes the intuitions much more vivid. The kind

overcrowded field, but I wanted a case that (a) is pretty realistic, and (b) doesn't involve the attribution (either to oneself or others) of a propositional attitude. In the example, X and Y are parents of a child, Z.

Y: This salad you bought is very good. Does it have nuts in it?

X: No. The nuttiness you're tasting is probably from the beans.

Y: Oh, so we could pack it for Z's lunch tomorrow.

X: Hang on, I better check about the nuts. Z's pre-school is very fussy about nuts. One of the children there might have an allergy, and it would be awful to get in trouble over her lunch.

Here's what I think is going on in that exchange. At  $t_2$  (I'll use  $t_i$  for the time of the  $i$ 'th utterance in the exchange), X believes that the salad has no nuts in it. Indeed, the one word sentence "No" expresses that belief. But by  $t_4$ , X has lost that belief. It would be fine to pack the salad for lunch if it has no nuts, but X isn't willing to do this for the simple reason that X no longer believes that it has no nuts. Moreover, this change of belief was, or at least could have been for all we've said so far, rational on X's part.

There's something a little puzzling about that. Jacob Ross and Mark Schroeder (forth) voice a common intuition when they say that beliefs should only change when new evidence comes in. Indeed, they use this intuition as a key argument against my view of belief. But X doesn't get any evidence that bears on the nuttiness of the salad. Yet X rationally changes beliefs. So I just conclude that sometimes we can change beliefs without new evidence coming in; sometimes our interests, broadly construed, change, and that is enough to change beliefs.

We'll come back to Ross and Schroeder's arguments in the next section, because first I want to concede something to the view that only evidence changes beliefs. That view is false, but there might be a true view in the area. And that's the view that only change in evidence can change *credences*. But that view only makes sense if there are such things as credences, and that's something that Richard Holton (forthcoming) has recently launched an intriguing argument against.

Holton's broader project is a much more sweeping attack on the Lockean thesis than I have proposed. Actually, it is a pair of more sweeping attacks. One of the pair is that the Lockeans identify something that exists, namely belief, with something that doesn't, namely high credence. I would not, could not, sign up for that critique. But I am much more sympathetic to the other side of his critique.

---

of cases we're considering play a big role in, *inter alia*, DeRose (1992); Cohen (1999) and Fantl and McGrath (2002).

Credences are, by their nature, exceedingly unstable. Whether an agent's credence that  $p$  is above or below 0.97, say, is liable to change according to any number of possible changes in evidence. But, at least if the agent is rational, beliefs are not so susceptible to change. Holton thinks that rational agents, or at least rational humans, frequently instantiate the following pattern. They form a belief that  $p$ , on excellent grounds. They later get some evidence that  $\neg p$ . The evidence is strong enough that, had they had it to begin with, they would have remained uncertain about  $p$ . But they do not decide to reopen the investigation into whether  $p$ . They hold on to their belief that  $p$ , the matter having been previously decided.

Such an attitude might look like unprincipled dogmatism. But it need not be, I think, as long as four criteria are met. (I think Holton agrees with these criteria.) One is that the agent's willingness to reopen the question of whether  $p$  must increase. She must be more willing, in the light of yet more evidence against  $p$ , to consider whether  $p$  is really true. A second is that, should the agent (irrationally) reopen the question of whether  $p$ , she should not use the fact that she previously closed that question as evidence. Once the genie is out of the box, only reasoning about  $p$  can get it back in. A third is that the costs of the inquiry must be high enough to warrant putting it off. If simply turning one's head fifteen degrees to the left will lead to acquiring evidence that definitively settles whether  $p$ , it is a little dogmatic to refuse to do so in the face of evidence against one's previously formed opinion that  $p$ . And finally, the costs of being wrong about  $p$  must not be too high. X, in our little dialogue above, would be terribly dogmatic if they didn't reopen the question of whether the salad had nuts in it, on being informed that this information was being used in a high stakes inquiry.

So beliefs should have a kind of resilience. Credences, if they exist, should not have this kind of resilience. So this suggests that a simple reduction of belief to credence, as the Lockeans suggest, cannot be right. You might worry that things are worse, that no reduction of belief to credence can be compatible with the difference in resilience between belief and credence. We'll return to that point, because first I want to look at Holton's stronger claim: that there are no such things as credences.

Holton acknowledges, as of course he must, that we have probabilistic truth-directed attitudes. We can imagine a person, call her Paula, who thinks it's likely that Richard III murdered his nephews, for instance. But Holton offers several reasons for thinking that in these probabilistic truth-directed attitudes, the probability goes in the content, not in the attitude. That is, we should interpret Paula as believing the probabilistic claim, Richard III probably murdered his nephews, and not as having some graded attitude towards the simple proposition *Richard*

*III murdered his nephews.* More precisely, Holton thinks we should understand Paula's *explicit* attitudes that way, and that independent of having reason to think that agents explicitly have probabilistic attitudes, there's no good way to make sense of the claim that they implicitly have probabilistic attitudes. So there's no such thing as credences, as usually understood. Or, at least, there's no good sense to be made of the claim that there are credences.

In response, I want to make six points.

1. Holton is right about cases like Paula's, and the possibility of iterating terms like *probably* provides independent support for this view.
2. Beliefs like the one Paula has are odd; they seem to have very strange truth conditions.
3. Our theory of mind needs some mechanism for explaining the relationship between confidence and action.
4. The 'explanatory gap' here could be filled by positing a binary attitude *is more confident that*.
5. This binary attitude can do all the work that graded attitudes were posited to do, and in a (historically sensitive) way saves the credences story.
6. Credences (or at least confidences) can have a key role within a Holton-like story about graded belief. They can both explain why agents reconsider some beliefs, and provide a standard of correctness for decisions to reconsider.

Let's take those in order.

I'm not going to rehearse Holton's argument for the 'content view': that in cases like Paula's the content of her attitude, and not the attitude itself, is probabilistic. But I do want to offer one extra consideration in its favour. (I'm indebted here to work in progress by my colleague Sarah Moss (2013), though I'm not sure she'd approve of this conclusion.) As well as Paula, we can imagine a person Pip who isn't sure that Paula is right, but thinks she's probably right. That is, Pip thinks that Richard III probably probably murdered his nephews. It's easy to make sense of Pip on the content view. Modalities in propositions iterate smoothly; that's what they are designed to do. But it's much harder to iterate attitudes. The possibility of cases like Pip suggests Holton must be right about Paula's case.

But Paula's case is odd. Beliefs have truth conditions. What are the truth conditions for Paula's belief? On the one hand, it seems they must be sensitive to her evidence. If she later acquires conclusive evidence that Richard III was framed, she won't think her earlier self had a false belief. But if we put the evidence into

the content of the belief, we get the strange result that her belief can't be preserved by uttering the same words to herself over again. That is, if the content of Paula's belief is *Given the evidence I have now, Richard III likely murdered his nephews*, she can't have the very same belief tomorrow by retaining the thought *Richard III likely murdered his nephews*. And she can't have a belief with the same content as anyone else by the two of them both thinking *Richard III likely murdered his nephews*. Those seem like unhappy conclusions, especially in the midst of a project that wants to emphasise the resiliency of belief. So perhaps we should say, following MacFarlane (2011), that the truth conditions of the belief are agent-relative. Or, if we're unhappy with the MacFarlane story, we might be pushed towards a kind of expressivism (perhaps a la Yalcin (2011)), which isn't quite like either the content view or the attitude view that Holton discusses. I'm personally partial to the relativist view, but I don't want to argue for that here, just note that the content view raises some interesting problems, and that natural solutions to them could in a way blur the boundaries between the content and attitude views.

As Holton notes in his discussion of Brutus, when our confidence in a proposition changes, our actions will change. Paula gets a little evidence that Richard III was framed, and her actions may change. Of course, not much of what we do in everyday life is sensitive to facts about English royal history, but there may be some effects. Maybe she'll be less inclined to speak up if the topic of the princes' murder comes up, or she'll take a slightly more jaundiced view of Shakespeare's play (compare Friend (2003).) Holton says that these falling confidences need not have all the precise structure of credences. In particular, they may not have the topology of the interval  $[0, 1]$ . But lots of credence lovers think that's too demanding. There's a long tradition of thinking that credences need not all be comparable.<sup>6</sup> What's important is that the relative confidences exist, and that they have a robust relationship to action.

There's an old fashioned way of doing this. The idea is implicit in Ramsey (1926), and made central in DeFinetti (1964). Take the binary attitude *is more confident that  $p$  than  $q$*  as primitive. As Holton notes, surface structure of our attitude reports suggest that this attitude, unlike the graded attitude of credence, is part of folk psychology. Lay down some constraints on this attitude. To get enough constraints that the binary relation determines a unique probability function, the constraints will have to be very tight. In particular, you'll need some kind of Archimedean principle, and a principle of universal comparability. Those aren't very plausible, especially the second. But even weaker constraints will get you something interesting. In particular, it isn't hard to lay down enough constraints that there is a unique set  $S$  of probability functions such that the agent is

---

<sup>6</sup>Notable members of the tradition include Levi (1974), Jeffrey (1983) and van Fraassen (1989).

more confident that  $p$  than  $q$  just in case  $\Pr(p) > \Pr(q)$  for all  $\Pr \in S$ . (For much more detail, see for instance Walley (1991).)

In that way, we can derive credences from the relative confidences of a reasonably coherent agent. But we can do with even less coherence than that I think. A throwaway remark from Ramsey (1929) provides a key clue. What is it to have credence  $2/3$  in  $p$ ? Don't say it's a betting disposition; mental states and behavioural dispositions aren't that tightly linked. Here's Ramsey's idea. To have credence  $2/3$  in  $p$  is to be exactly as confident in  $p$  as in  $q \vee r$ , where  $q$ ,  $r$  and  $s$  are taken to be exclusive and exhaustive, and one has equal confidence in all three. It's easy to see how to extend that to a definition of credence  $m/n$  for any integer  $m, n$ . It's a little trickier to say precisely what, say, credence  $1/\pi$  is, but rational credences are probably credences enough to explain action. And just like that, we have a way of talking about credences, i.e., graded attitudes, without positing anything more than a binary attitude *more confident than*.

Perhaps Holton could argue that we only have unary attitudes, not binary attitudes like *more confident than*. If Maury is more confident that Oswald shot JFK than that Richard III murdered his nephews, that means he really believes the proposition *It is more likely that Oswald shot JFK than that Richard III murdered his nephews*. But such a view seems forced at best, and isn't motivated by Holton's other arguments for the 'content view'. This attitude of *more confident than* isn't iterable. It isn't subject to the particular kind of reasoning errors that Holton takes to be evidence for the content view in the probabilistic case. It is an attitude we ordinarily report as a binary attitude in normal speech. In short, it looks like a genuine binary attitude.

Given that the binary attitude exists, and that we can define numerical (at least rational) credences in terms of it, I'd say that's enough to say that credences exist. In a sense, credences will be epiphenomenal. What does the explanatory work is the binary relation *more confident than*. Maury might stay away from a showing of *Richard III* because he is less confident that it is historically accurate than he used to be. We can work out from Maury's other relative confidences what his credence in Richard III's guilt is and was. Or, at least, we can work out bounds on these. But those numbers aren't in a fundamental sense explanatory, and neither are the complicated sets of relative confidences that constitute the numbers. What's *really* explanatory are relative confidences. But it's a harmless enough mode of speech to talk as if credences are explanatory; they are easier to talk about than the underlying relative confidences.



### 3 The Power of Theoretical Interests

So I think we should accept that credences exist. And we can just about reduce beliefs to credences. In previous work I argued that we could do such a reduction. I'm not altogether sure whether the amendments to that view I'm proposing here means it no longer should count as a reductive view; we'll come back to that question in the conclusion.

The view I defended in previous work is that the reduction comes through the relationship between conditional and unconditional attitudes. Very roughly, to believe that  $p$  is simply to have the same attitudes, towards all salient questions, unconditionally as you have conditional on  $p$ . In a syrupy slogan, belief means never having to say you've conditionalised. For reasons I mentioned in section 1, I now think that was inaccurate; I should have said that belief means never having to say you've updated, or at least that you've updated your view on any salient question.

The restriction to salient questions is important. Consider any  $p$  that I normally take for granted, but such that I wouldn't bet on it at insane odds. I prefer declining such a bet to taking it. But conditional on  $p$ , I prefer taking the bet. So that means I don't believe any such  $p$ . But just about any  $p$  satisfies that description, for at least some 'insane' odds. So I believe almost nothing. That would be a *reductio* of the position. I respond by saying that the choice of whether to take an insane bet is not normally salient.

But now there's a worry that I've let in too much. For many  $p$ , there is no salient decision that they even bear on. What I would do conditional on  $p$ , conditional on  $\neg p$ , and unconditionally is exactly the same, over the space of salient choices. (And this isn't a case where updating and conditionalising come apart; I'll leave this proviso mostly implicit from now on.) So with the restriction in place, I believe  $p$  and  $\neg p$ . That seems like a *reductio* of the view too. I probably do have inconsistent beliefs, but not in virtue of  $p$  being irrelevant to me right now. I've changed my mind a little about what the right way to avoid this problem is, in part because of some arguments by Jacob Ross and Mark Schroeder.

They have what looks like, on the surface, a rather different view to mine. They say that to believe  $p$  is to have a **default reasoning disposition** to use  $p$  in reasoning. Here's how they describe their own view.

What we should expect, therefore, is that for some propositions we would have a *defeasible* or *default* disposition to treat them as true in our reasoning—a disposition that can be overridden under circumstances where the cost of mistakenly acting as if these propositions

are true is particularly salient. And this expectation is confirmed by our experience. We do indeed seem to treat some uncertain propositions as true in our reasoning; we do indeed seem to treat them as true automatically, without first weighing the costs and benefits of so treating them; and yet in contexts such as High where the costs of mistakenly treating them as true is salient, our natural tendency to treat these propositions as true often seems to be overridden, and instead we treat them as merely probable.

But if we concede that we have such defeasible dispositions to treat particular propositions as true in our reasoning, then a hypothesis naturally arises, namely, that beliefs consist in or involve such dispositions. More precisely, at least part of the functional role of belief is that believing that  $p$  defeasibly disposes the believer to treat  $p$  as true in her reasoning. Let us call this hypothesis the *reasoning disposition account* of belief.

There are, relative to what I'm interested in, three striking characteristics of Ross and Schroeder's view.

1. Whether you believe  $p$  is sensitive to how you reason; that is, your theoretical interests matter.
2. How you would reason about some questions that are not live is relevant to whether you believe  $p$ .
3. Dispositions can be masked, so you can believe  $p$  even though you don't actually use  $p$  in reasoning now.

I think they take all three of these points to be reasons to favour their view over mine. As I see it, we agree on point 1 (and I always had the resources to agree with them), I can accommodate point 2 with a modification to my theory, and point 3 is a cost of their theory, not a benefit. Let's take those points in order.

There are lots of reasons to dislike what Ross and Schroeder call *Pragmatic Credal Reductionism* (PCR). This is, more or less, the view that the salient questions, in the sense relevant above, are just those which are practically relevant to the agent. So to believe  $p$  just is to have the same attitude towards all practically relevant questions unconditionally as conditional on  $p$ . There are at least three reasons to resist this view.

One reason comes from the discussions of Ned Block's example Blockhead (Block, 1978). As Braddon-Mitchell and Jackson point out, the lesson to take from that example is that beliefs are constituted in part by their relations to other

mental states (Braddon-Mitchell and Jackson, 2007, 114ff). There's a quick attempted refutation of PCR via the Blockhead case which doesn't quite work. We might worry that if all that matters to belief given PCR is how it relates to action, PCR will have the implausible consequence that Blockhead has a rich set of beliefs. That isn't right; PCR is compatible with the view that Blockhead doesn't have credences, and hence doesn't have credences that constitute beliefs. But the Blockhead examples value isn't exhausted by its use in quick refutations.<sup>7</sup> The lesson is that beliefs are, by their nature, interactive. It seems to me that PCR doesn't really appreciate that lesson.

Another reason comes from recent work by Jessica Brown (forthcoming). Compare these two situations.

1. *S* is in circumstances *C*, and has to decide whether to do *X*.
2. *S* is in completely different circumstances to *C*, but is seriously engaged in planning for future contingencies. She's currently trying to decide whether in circumstances *C* to do *X*.

Intuitively, *S* can bring exactly the same evidence, knowledge and beliefs to bear on the two problems. If *C* is a particularly high stakes situation, say it is a situation where one has to decide what to feed someone with a severe peanut allergy, then a lot of things that can ordinarily be taken for granted cannot, in this case, be taken for granted. And that's true whether *S* is actually in *C*, or she is just planning for the possibility that she finds herself in *C*.

So I conclude that both practical and theoretical interests matter for what we can take for granted in inquiry. The things we can take for granted into a theoretical inquiry into what to do in high stakes contexts as restricted, just as they are when we are in a high stakes context, and must make a practical decision. Since the latter restriction on what we can take for granted is explained by (and possibly constituted by) a restriction on what we actually believe in those contexts, we should similarly conclude that agents simply believe less when they are reasoning about high stakes contexts, whatever their actual context.

A third reason to dislike PCR comes from the 'Renzi' example in Ross and Schroeder's paper. I'll run through a somewhat more abstract version of the case, because I don't think the details are particularly important. Start with a standard decision problem. The agent knows that *X* is better to do if *p*, and *Y* is better to do if  $\neg p$ . The agent should then go through calculating the relative gains to

---

<sup>7</sup>The point I'm making here is relevant I think to recent debates about the proper way to formalise counterexamples in philosophy, as in (Williamson, 2007b; Ichikawa and Jarvis, 2009; Malmgren, 2011). I worry that too much of that debate is focussed on the role that examples play in one-step refutations. But there's more, much more, to a good example than that.

doing X or Y in the situations they are better, and the probability of  $p$ . But the agent imagined doesn't do that. Rather, the agent divides the possibility space in four, taking the salient possibilities to be  $p \wedge q$ ,  $p \wedge \neg q$ ,  $\neg p \wedge q$  and  $\neg p \wedge \neg q$ , and then calculates the expected utility of X and Y accordingly. This is a bad bit of reasoning on the agent's part. In the cases we are interested in,  $q$  is exceedingly likely. Moreover, the expected utility of each act doesn't change a lot depending on  $q$ 's truth value. So it is fairly obvious that we'll end up making the same decision whether we take the 'small worlds' in our decision model to be just the world where  $p$ , and the world where  $\neg p$ , or the four worlds this agent uses. But the agent does use these four, and the question is what to say about them.

Ross and Schroeder say that such an agent should not be counted as believing that  $q$ . If they are consciously calculating the probability that  $q$ , and taking  $\neg q$  possibilities into account when calculating expected utilities, they regard  $q$  as an open question. And regarding  $q$  as open in this way is incompatible with believing it. I agree with all this.

They also think that PCR implies that the agent *does* believe  $q$ . The reason is that conditionalising on  $q$  doesn't change the agent's beliefs about any practical question. I think that's right too, at least on a natural understanding of what 'practical' is.

My response to all these worries is to say that whether someone believes that  $p$  depends not just on how conditionalising (or more generally updating) on  $p$  would affect someone's action, but on how it would affect their reasoning. That is, just as we learned from the Blockhead example, to believe that  $p$  requires having a mental state that is connected to the rest of one's cognitive life in roughly the way a belief that  $p$  should be connected. Let's go through both the last two cases to see how this works on my theory.

One of the things that happens when the stakes go up is that conditionalising on very probable things can change the outcome of interesting decisions. Make the probability that some nice food is peanut-free be high, but short of one. Conditional on it being peanut-free, it's a good thing to give to a peanut-allergic guest. But unconditionally, it's a bad thing to give to such a guest, because the niceness of the food doesn't outweigh the risk of killing them. And that's true whether the guest is actually there, or you're just thinking about what to do should such a guest arrive in the future. In general, the same questions will be relevant whether you're in  $C$  trying to decide whether to do  $X$ , or simply trying to decide whether to  $X$  in  $C$ . In one case they will be practically relevant questions, in the other they will be theoretically relevant questions. But this feels a lot like a distinction without a difference, since the agent should have similar beliefs in the two cases.

The same response works for Ross and Schroeder's case. The agent was trying to work out the expected utility of X and Y by working out the utility of each action in each of four 'small worlds', then working out the probability of each of these. Conditional on  $q$ , the probability of two of them ( $p \wedge \neg q, \neg p \wedge \neg q$ ), will be 0. Unconditionally, this probability won't be 0. So the agent has a different view on some question they have taken an interest in unconditionally to their view conditional on  $q$ . So they don't believe  $q$ . The agent shouldn't care about that question, and conditional on each question they should care about, they have the same attitude unconditionally and conditional on  $q$ . But they do care about these probabilistic questions, so they don't believe  $q$ . (In (Weatherson, 2005) I said that to justifiably believe  $q$  was to have a justified credence in  $q$  that was sufficiently high to count as a belief. The considerations of the last two sentences puts some pressure on that reductive theory of justification for beliefs.)

So I think that Ross and Schroeder and I agree on point 1; something beyond practical interests is relevant to belief.

They have another case that I think does suggest a needed revision to my theory. I'm going to modify their case a little to change the focus a little, and to avoid puzzles about vagueness. (What follows is a version of their example about Dalí's moustache, purged of any worries about vagueness, and without the focus on consistency. I don't think the problem they true to press on me, that my theory allows excessive inconsistency of belief among rational agents, really sticks. Everyone will have to make qualifications to consistency to deal with the preface paradox, and for reasons I went over in (Weatherson, 2005), I think the qualifications I make are the best ones to make.)

Let  $D$  be the proposition that the number of games the Detroit Tigers won in 1976 (in the MLB regular season) is not a multiple of 3. At most times,  $D$  is completely irrelevant to anything I care about, either practically or theoretically. My attitudes towards any relevant question are the same unconditionally as conditional on  $D$ . So there's a worry that I'll count as believing  $D$ , and believing  $\neg D$ , by default.

In earlier work, I added a clause meant to help with cases like this. I said that for determining whether an agent believes that  $p$ , we should treat the question of whether  $p$ 's probability is above or below 0.5 as salient, even if the agent doesn't care about it. Obviously this won't help with this particular case. The probability of  $D$  is around  $2/3$ , and is certainly above 0.5. My 'fix' avoids the consequence that I implausibly count as believing  $\neg D$ . But I still count, almost as implausibly, as believing  $D$ . This needs to be fixed.

Here's my proposed change. For an agent to count as believing  $p$ , it must be possible for  $p$  to do some work for them in reasoning. Here's what I mean by work. Consider a very abstract set up of a decision problem, as follows.

|   | $p$ | $q$ |
|---|-----|-----|
| X | 4   | 1   |
| Y | 3   | 2   |

That table encodes a lot of information. It encodes that  $p \vee q$  is true; otherwise there are some columns missing. It encodes that the only live choices are X or Y; otherwise there are rows missing. It encodes that doing X is better than doing Y if  $p$ , and worse if  $q$ .

For any agent, and any decision problem, there is a table like this that they would be disposed to use to resolve that problem. Or, perhaps, there are a series of tables and there is no fact about which of them they would be most disposed to use.

Given all that terminology, here's my extra constraint on belief. To believe that  $p$ , there must be some decision problem such that some table the agent would be disposed to use to solve it encodes that  $p$ . If there is no such problem, the agent does not believe that  $p$ . For anything that I intuitively believe, this is an easy condition to satisfy. Let the problem be whether to take a bet that pays 1 if  $p$ , and loses 1 otherwise. Here's the table I'd be disposed to use to solve the problem.

|             | $p$ |
|-------------|-----|
| Take bet    | 1   |
| Decline bet | 0   |

This table encodes that  $p$ , so it is sufficient to count as believing that  $p$ . And it doesn't matter that this bet isn't on the table. I'm disposed to use this table, so that's all that matters.

But might there be problems in the other direction. What about an agent who, if offered such a bet on  $D$ , would use such a simple table? I simply say that they believe that  $D$ . I would not use any such table. I'd use this table.

|             | $D$ | $\neg D$ |
|-------------|-----|----------|
| Take bet    | 1   | -1       |
| Decline bet | 0   | 0        |

Now given the probability of  $D$ , I'd still end up taking the bet; it has an expected return of  $2/3$ . (Well, actually I'd probably decline the bet because being offered the bet would change the probability of  $D$  for reasons made clear in (Runyon, 1992, 14–15). But that hardly undermines the point I'm making.) But this isn't some analytic fact about me, or even I think some respect in which I'm obeying the dictates of rationality. It's simply a fact that I wouldn't take  $D$  for granted in any inquiry. And that's what my non-belief that  $D$  consists in.

This way of responding to the Tigers example helps respond to a nice observation that Ross and Schroeder make about correctness. A belief that  $p$  is, in some sense, *incorrect* if  $\neg p$ . It isn't altogether clear how to capture this sense given a simple reduction of beliefs to credences. I propose to capture it using this idea that decision tables encode propositions. A table is incorrect if it encodes something that's false. To believe something is, *inter alia*, to be disposed to use a table that encodes it. So if it is false, it involves a disposition to do something incorrect.

It also helps capture Holton's observation that beliefs should be resilient. If someone is disposed to use decision tables that encode that  $p$ , that disposition should be fairly resilient. And to the extent that it is resilient, they will satisfy all the other clauses in my preferred account of belief. So anyone who believes  $p$  should have a resilient belief that  $p$ .

The last point is where I think my biggest disagreement with Ross and Schroeder lies. They think it is very important that a theory of belief vindicate a principle they call **Stability**.

**Stability:** A fully rational agent does not change her beliefs purely in virtue of an evidentially irrelevant change in her credences or preferences. (20)

Here's the kind of case that is meant to motivate Stability, and show that views like mine are in tension with it.

Suppose Stella is extremely confident that steel is stronger than Styrofoam, but she's not so confident that she'd bet her life on this proposition for the prospect of winning a penny. PCR implies, implausibly, that if Stella were offered such a bet, she'd cease to believe that steel is stronger than Styrofoam, since her credence would cease to rationalize acting as if this proposition is true. (22)

Ross and Schroeder's own view is that if Stella has a defeasible disposition to treat as true the proposition that steel is stronger than Styrofoam, that's enough for her to believe it. And that can be true if the disposition is not only defeasible, but actually defeated in the circumstances Stella is in. This all strikes me as just as implausible as the failure of Stability. Let's go over its costs.

The following propositions are clearly not mutually consistent, so one of them must be given up. We're assuming that Stella is facing, and knows she is facing, a bet that pays a penny if steel is stronger than Styrofoam, and costs her life if steel is not stronger than Styrofoam.

1. Stella believes that steel is stronger than Styrofoam.
2. Stella believes that if steel is stronger than Styrofoam, she'll win a penny and lose nothing by taking the bet.
3. If 1 and 2 are true, and Stella considers the question of whether she'll win a penny and lose nothing by taking the bet, she'll believe that she'll win a penny and lose nothing by taking the bet.
4. Stella prefers winning a penny and losing nothing to getting nothing.
5. If Stella believes that she'll win a penny and lose nothing by taking the bet, and prefers winning a penny and losing nothing to getting nothing, she'll take the bet.
6. Stella won't take the bet.

It's part of the setup of the problem that 2 and 4 are true. And it's common ground that 6 is true, at least assuming that Stella is rational. So we're left with 1, 3 and 5 as the possible candidates for falsehood.

Ross and Schroeder say that it's implausible to reject 1. After all, Stella believed it a few minutes ago, and hasn't received any evidence to the contrary. And I guess rejecting 1 isn't the most intuitive philosophical conclusion I've ever drawn. But compare the alternatives!

If we reject 3, we must say that Stella will simply refuse to infer  $r$  from  $p$ ,  $q$  and  $(p \wedge q) \rightarrow r$ . Now it is notoriously hard to come up with a general principle for closure of beliefs. But it is hard to see why this particular instance would fail. And in any case, it's hard to see why Stella wouldn't have a general, defeasible, disposition to conclude  $r$  in this case, so by Ross and Schroeder's own lights, it seems 3 should be acceptable.

That leaves 5. It seems on Ross and Schroeder's view, Stella simply must violate a very basic principle of means-end reasoning. She desires something, she believes that taking the bet will get that thing, and come with no added costs. Yet, she refuses to take the bet. And she's rational to do so! At this stage, I think I've lost what's meant to be belief-like about their notion of belief. I certainly think



attributing this kind of practical incoherence to Stella is much less plausible than attributing a failure of Stability to her.

Put another way, I don't think presenting Stability on its own as a desideratum of a theory is exactly playing fair. The salient question isn't whether we should accept or reject Stability. The salient question is whether giving up Stability is a fair price to pay for saving basic tenets of means-end rationality. And I think that it is. Perhaps there will be some way of understanding cases like Stella's so that we don't have to choose between theories of belief that violate Stability constraints, and theories of belief that violate coherence constraints. But I don't see one on offer, and I'm not sure what such a theory could look like.

I have one more argument against Stability, but it does rest on somewhat contentious premises. There's often a difference between the best *methodology* in an area, and the correct *epistemology* of that area. When that happens, it's possible that there is a good methodological rule saying that if such-and-such happens, re-open a certain inquiry. But that rule need not be epistemologically significant. That is, it need not be the case that the happening of such-and-such provides evidence against the conclusion of the inquiry. It just provides a reason that a good researcher will re-open the inquiry. And, as we've stated above, an open inquiry is incompatible with belief.

Here's one way that might happen. Like other non-conciliationists about disagreement, e.g., (Kelly, 2010), I hold that disagreement by peers with the same evidence as you doesn't provide *evidence* that you are wrong. But it might provide an excellent reason to re-open an inquiry. We shouldn't draw conclusions about the methodological significance of disagreement from the epistemology of disagreement. So learning that your peers all disagree with a conclusion might be a reason to re-open inquiry into that conclusion, and hence lose belief in the conclusion, without providing evidence that the conclusion is false. This example rests on a very contentious claim about the epistemology of disagreement. But any gap that opens up between methodology and epistemology will allow such an example to be constructed, and hence provide an independent reason to reject Stability.

## 4 Conclusion

You might well worry that the view here is too *complex* to really be a theory of belief. Belief is a simple state; why all the epicycles? This is a good question, and I'm not sure I have a sufficiently good answer to it.

At heart, the theory I've offered here is simple. To believe  $p$  is to take  $p$  for granted, to take it as given, to take it as a settled question. But one doesn't take

a question as settled in a vacuum. I will take some questions as settled in some circumstances and not others. It's here that the complexities enter in.

To believe  $p$ , it isn't necessary that we take it as settled in all contexts. That would mean that anything one believes one would bet on at any odds. But it isn't sufficient to take it as settled in some context or other. If I'm facing a tricky bet on  $p$ , the fact that I'd take  $p$  as settled in some other context doesn't mean that I believe  $p$ . After all, I might even decline the bet, although I desire the reward for winning the bet, and believe that if  $p$  I will win. And we can't just focus on the actual circumstances. Five minutes ago, I neither took it as settled or as open that the Cubs haven't won the World Series for quite a while. I simply wasn't thinking about that proposition, and didn't really take it to be one thing or another.

This is why things get so complex. To believe  $p$  is to hold a fairly simple attitude towards  $p$  in some relevant circumstances. But which circumstances? That's what's hard to say, and it's why the theory is so messy. And I think we have an argument that it must be a little hard to say, namely an argument by exhaustion of all the possible simple things to say. The previous paragraph starts such an argument.

I'd be a little surprised if the account here is the best or last word on the matter though. It does feel a little disjunctive, as if there is a simpler reduction to be had. But I think it's better than what came before, so I'm putting it forward.

The previous version of the theory I put forward was clearly reductive; beliefs were reduced to credences and preferences. This version is not quite as clearly reductive. Which decision tables the agent is disposed to use, and which propositions those tables encode, are not obviously facts about credences and preferences. So it feels like I've given up on the reductive project.

I'm not altogether happy about this; reduction is a good aim to have. But if reduction of belief to other states fails, I'd think this kind of reason is why it is going to fail. Facts about how an agent conceptualises a problem, how she sets up the decision table, are distinct from facts about which values she writes into the table. This is the deepest reason why the Lockean theory is false. Belief is not the difference between one column in the decision table getting probability 0.98 rather than 0.97; it is the difference between one column being excluded rather than included. If that difference can't be accounted for in terms of actual credences and preferences, the reductionist project will fail.

## References

- Aumann, Robert J. 1999. "Interactive Epistemology I: Knowledge." *International Journal of Game Theory* 28:263–300.

- Binmore, Ken. 2007. *Playing for Real: A Text on Game Theory*. Oxford: Oxford University Press.
- Block, Ned. 1978. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9:261–325.
- Braddon-Mitchell, David and Jackson, Frank. 2007. *The Philosophy of Mind and Cognition, second edition*. Malden, MA: Blackwell.
- Brown, Jessica. forthcoming. "Impurism, Practical Reasoning and the Threshold Problem." *Noûs*, doi:10.1111/nous.12008.
- Christensen, David. 2005. *Putting Logic in Its Place*. Oxford: Oxford University Press.
- Cohen, Stewart. 1999. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives* 13:57–89, doi:10.1111/0029-4624.33.s13.3.
- DeFinetti, Bruno. 1964. "Foresight: Its Logical Laws, Its Subjective Sources." In Henry E. Kyburg and Howard E. Smokler (eds.), *Studies in Subjective Probability*, 93–156. New York: Wiley. First published as "La prévision : ses lois logiques, ses sources subjectives", in *Annals de l'institut Henri Poincaré*, vol 7, number 1, 1937, pp 1-68.
- DeRose, Keith. 1992. "Contextualism and Knowledge Attributions." *Philosophy and Phenomenological Research* 52:513–529s.
- Dixit, Avinash K. and Skeath, Susan. 2004. *Games of Strategy*. New York: W. W. Norton & Company, second edition.
- Egan, Andy and Weatherson, Brian (eds.). 2011. *Epistemic Modality*. Oxford: Oxford University Press.
- Fantl, Jeremy and McGrath, Matthew. 2002. "Evidence, Pragmatics, and Justification." *Philosophical Review* 111:67–94.
- van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.
- Friend, Stacie. 2003. "How I Really Feel about JFK." In Matthew Kieran and Dominic McIver Lopes (eds.), *Imagination, Philosophy and the Arts*, London, 35–53. Routledge.
- Gillies, Anthony S. 2010. "Iffiness." *Semantics and Pragmatics* 3:1–42, doi:10.3765/sp.3.4.

- Holton, Richard. forthcoming. "Intention as a Model for Belief." In Manuel Vargas and Gideon Yaffe (eds.), *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*. Oxford: Oxford University Press.
- Ichikawa, Jonathan and Jarvis, Benjamin. 2009. "Thought-experiment intuitions and truth in fiction." *Philosophical Studies* 142:221–246, doi:10.1007/s11098-007-9184-y.
- Jeffrey, Richard. 1983. "Bayesianism with a Human Face." In J. Earman (ed.) (ed.), *Testing Scientific Theories*. Minneapolis: University of Minnesota Press.
- Kelly, Thomas. 2010. "Peer disagreement and higher order evidence." In Ted Warfield and Richard Feldman (eds.), *Disagreement*, 111–174. Oxford: Oxford University Press.
- Kohlberg, Elon and Mertens, Jean-Francois. 1986. "On the Strategic Stability of Equilibria." *Econometrica* 54:1003–1037.
- Levi, Isaac. 1974. "On Indeterminate Probabilities." *Journal of Philosophy* 71:391–418.
- MacFarlane, John. 2011. "Epistemic Modals are Assessment-Sensitive." In Egan and Weatherson (2011), 144–178.
- Malmgren, Anna-Sara. 2011. "Rationalism and the Content of Intuitive Judgments." *Mind* 120:263–327, doi:10.1093/mind/fzr039.
- Moss, Sarah. 2013. "On the Pragmatics of Epistemic Modals." Manuscript.
- Ramsey, Frank. 1926. "Truth and Probability." In Ramsey (1990), 52–94.
- . 1929. "Probability and Partial Belief." In Ramsey (1990), 95–96.
- . 1990. *Philosophical Papers*. Cambridge: Cambridge University Press.
- Ross, Jacob and Schroeder, Mark. fort. "Belief, Credence, and Pragmatic Encroachment." *Philosophy and Phenomenological Research* , doi:10.1111/j.1933-1592.2011.00552.xForthcoming.
- Runyon, Damon. 1992. *Guys & Dolls: The stories of Damon Runyon*. New York: Penguin.
- Sorensen, Roy A. 1988. *Blindspots*. Oxford: Clarendon Press.

- Stalnaker, Robert. 1994. "On the evaluation of solution concepts." *Theory and Decision* 37:49–73, doi:10.1007/BF01079205.
- . 1998. "Belief revision in games: forward and backward induction." *Mathematical Social Sciences* 36:31 – 56. ISSN 0165-4896, doi:10.1016/S0165-4896(98)00007-9.
- . 1999. "Extensive and strategic forms: Games and models for games." *Research in Economics* 53:293 – 319. ISSN 1090-9443, doi:10.1006/reec.1999.0200.
- Stalnaker, Robert C. 1996. "Knowledge, Belief and Counterfactual Reasoning in Games." *Economics and Philosophy* 12:133–163, doi:10.1017/S0266267100004132.
- Walley, Peter. 1991. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman & Hall.
- Weatherson, Brian. 2005. "Can We Do Without Pragmatic Encroachment?" *Philosophical Perspectives* 19:417–443, doi:10.1111/j.1520-8583.2005.00068.x.
- . 2011. "Defending Interest-Relative Invariantism." *Logos & Episteme* 2:591–609.
- . 2012a. "Games and the Reason-Knowledge Principle." *The Reasoner* 6:6–8.
- . 2012b. "Knowledge, Bets and Interests." In Jessican Brown and Mikkel Gerken (eds.), *Knowledge Ascriptions*, 75–103. Oxford: Oxford University Press.
- White, Roger. 2005. "Epistemic permissiveness." *Philosophical Perspectives* 19:445–459.
- Williamson, Timothy. 2007a. "How probable is an infinite sequence of heads?" *Analysis* 67:173–180, doi:10.1111/j.1467-8284.2007.00671.x.
- . 2007b. *The Philosophy of Philosophy*. Blackwell Pub. Ltd.
- . 2012. "Models of Improbable Knowing." Presented at the 2012 CSMN/Arché epistemology conference, Lofoten Islands, May 2012.
- Yalcin, Seth. 2011. "Nonfactualism about Epistemic Modality." In Egan and Weatherson (2011), 295–332.