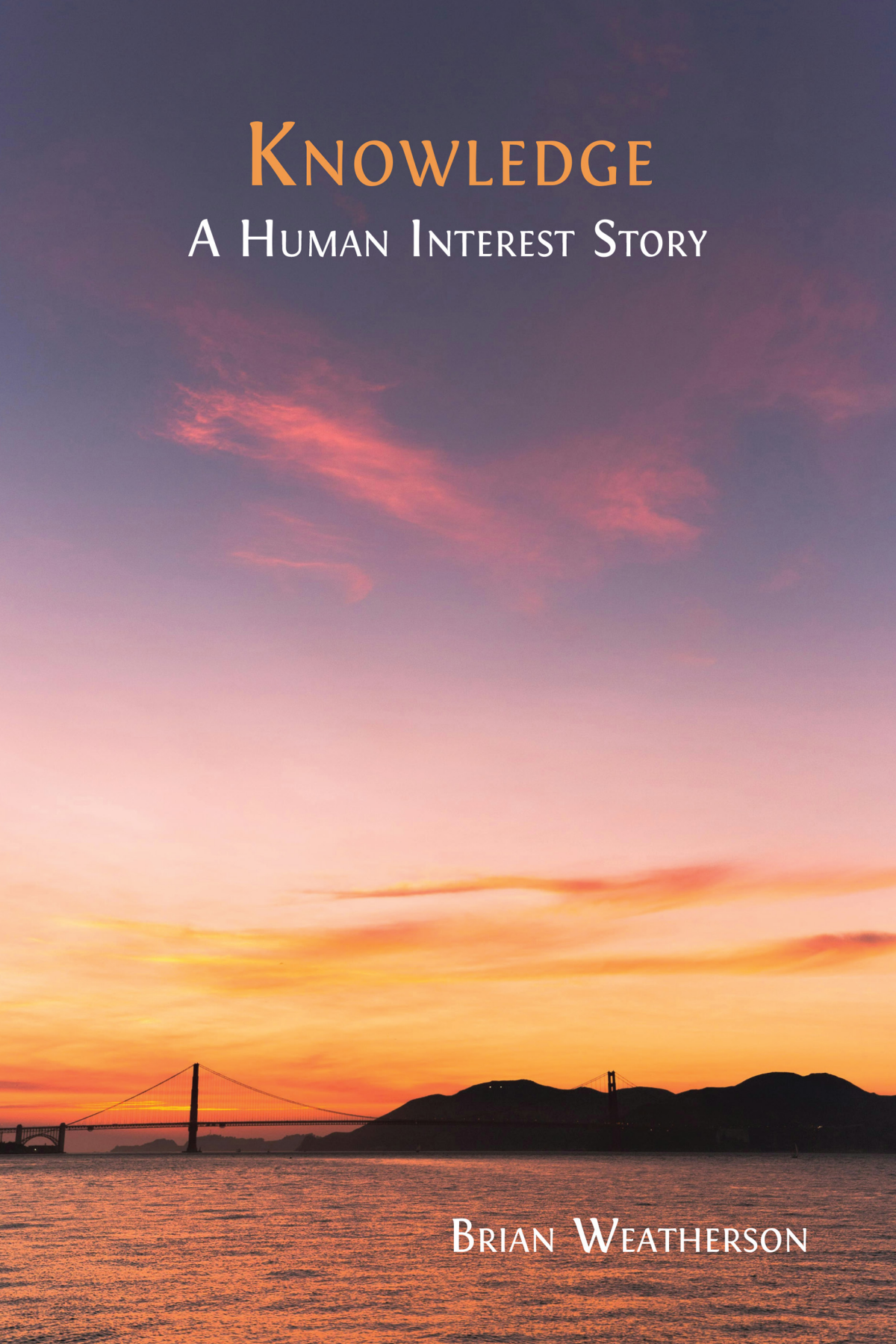
Knowledge

A Human Interest Story



Brian Weatherson

https://www.openbookpublishers.com

ISBN Paperback: 978-1-80511-394-2

ISBN Hardback: 978-1-80511-395-9

ISBN Digital (PDF): 978-1-80511-396-6

ISBN Digital eBook (EPUB): 978-1-80511-397-3

ISBN HTML: 978-1-80511-398-0

DOI: 10.11647/OBP.0425

Cover image: Golden Gate Bridge, San Francisco, California. Photo by Tj Kolesnik, https://unsplash.com/photos/silhouette-photography-of-mountain-beside-suspension-bridgeduring-golden-hour-Vnz7o7LeuDs.

Cover design: Jeevanjot Kaur Nagpal

8. Rationality

This chapter is about rational belief. My version of IRT allows a new kind of gap between rational true belief and knowledge, and I'll argue we should treat this as a philosophical discovery, not a refutation of the view. Then I'll present two arguments for the possibility of rationally having credence 1 in a proposition without believing it. The first is due to Timothy Williamson; the second is new. These arguments refute two claims about the relationship between belief and credence. One is a descriptive claim: to believe p just is to have credence in p at or above some threshold. The other is a normative claim: one rationally believes p just in case one rationally has credence in it at or above some threshold. Even if those two arguments concerning belief and credence 1 don't work, and rational credence 1 does entail rational belief, there are independent arguments against the descriptive and normative claims if the 'threshold' in them is non-maximal. I'll end the chapter by noting how the view of rational belief that comes out of IRT is immune to the problems associated with understanding belief in terms of a credal threshold.

8.1 Atomism about Rational Belief

In Chapter 3 I suggested that the following two conditions were individually necessary for belief that p, and suggested they might be jointly sufficient.[1]

1. In some possible decision problem, p is taken for granted.
2. For every question the agent is interested in, the agent answers the question the same way (i.e., giving the same

1 This section is based on Weatherson (2012).

 https://doi.org/10.11647/OBP.0425.08

answer for the same reasons) whether the question is asked unconditionally or conditional on p.

At this point one might think that offering a theory of rational belief would be easy. It is rational to believe p just in case it is rational to satisfy these conditions. Unfortunately, this nice thought can't be right. It can be irrational to satisfy these conditions while rationally believing p.

Coraline is like Anisa and Chamari, in that she has read a reliable book saying that the Battle of Agincourt was in 1415. And she now believes that the Battle of Agincourt was indeed in 1415, for the very good reason that she read it in a reliable book.

In front of her is a sealed envelope, and inside the envelope a number is written on a slip of paper. Let X denote that number, non-rigidly. (So when I say Coraline believes $X = x$, it means she believes that the number written on the slip of paper is x, where x rigidly denotes some number.) Coraline is offered the following bet:

- If she declines the bet, nothing happens.
- If she accepts the bet, and the Battle of Agincourt was in 1415, she wins $1.
- If she accepts the bet, and the Battle of Agincourt was not in 1415, she loses X dollars.

For some reason, Coraline is convinced that $X = 10$. This is very strange, since she was shown the slip of paper just a few minutes ago, and it clearly showed that $X = 10^9$. Coraline wouldn't bet on when the Battle of Agincourt was at odds of a billion to one. But she would take that bet at ten to one, which is what she thinks she is faced with. Indeed, she doesn't even conceptualise it as a bet; it's a free dollar she thinks. Right now, she is disposed to treat the date of the battle as a given. She is disposed to lose this disposition should a very long-odds bet appear to depend on it. But she doesn't believe she is facing such a bet.

So Coraline accepts the bet; she thinks it is a free dollar. Since that's when the battle took place, she wins the dollar. All's well that ends well. But it really was a wildly irrational bet to take. You shouldn't bet at those odds on something you remember from a history book. Neither memory nor history books are that reliable. Coraline was not rational to treat the questions *Should I take this bet?*, and *Conditional on the Battle of Agincourt*

being in 1415, should I take this bet? the same way. Her treating them the same way was fortunate—she won a dollar—but irrational.

Yet it seems odd to say that Coraline's belief about the Battle of Agincourt was irrational. What was irrational was her belief about the envelope, not her belief about the battle. To say that a particular disposition was irrational is to make a holistic assessment of the person with the disposition. But whether a belief is rational or not is, relatively speaking, atomistic.

That suggests the following condition on rational belief: S's belief that p is irrational if

1. S irrationally has one of the dispositions that is characteristic of belief that p; and
2. What explains S having a disposition that is irrational in that way is her attitudes towards p, not (solely) her attitudes towards other propositions, or her skills in practical reasoning.

Intuitively, Coraline's irrational acceptance of the belief is explained by her (irrational) belief about what's in the envelope, not her (rational) belief about the Battle of Agincourt. We can take the relevant notion of explanation as a primitive if we like; it's in no worse philosophical shape than other notions we take as a primitive. But it is possible to spell it out a little more.

Coraline has a pattern of irrational dispositions related to the envelope. If you offer her \$50 or X dollars, she'll take the \$50. Alternatively, if you change the bet so it isn't about Agincourt, but is instead about any other thing for which she has excellent but not quite conclusive evidence, she'll still take the bet. On the other hand, she does not have a pattern of irrational dispositions related to the Battle of Agincourt. She has this one, but if you change the payouts so they are not related to this particular envelope, then for all we have said so far, she won't do anything irrational.

That difference in patterns matters. We know that it's the beliefs about the envelope, and not the beliefs about the battle, that are explanatory because of this pattern. We could try and create a reductive analysis of explanation in clause 2 using facts about patterns, like the way David Lewis tries to create a reductive analysis of causation using

similar facts about patterns in "Causation as Influence" (Lewis, 2004). But doing so would invariably run up against edge cases that would be more trouble to resolve than they are worth. There are ever so many ways in which someone could have an irrational disposition about any particular case. We can imagine Coraline having a rational belief about the envelope, but still taking the bet because of any of the following reasons:

- It has been her life goal to lose a billion dollars in a day, so taking the bet strictly dominates not taking it.
- She believes (irrationally) that anyone who loses a billion dollars in a day goes to heaven, and she (rationally) values heaven above any monetary amount.
- She consistently makes reasoning errors about billions, so the prospect of losing a billion dollars rarely triggers an awareness that she should reconsider things she normally takes for granted.

The last one of these is especially interesting. The picture of rational agency I'm working with here owes a lot to the notion of epistemic vigilance, as developed by Dan Sperber and colleagues (Sperber et al., 2010). The rational agent will have all these beliefs in their head that they will drop when the costs of being wrong about them are too high, or the costs of re-opening inquiry into them are too low. They can't reason, at least in any conscious way, about whether to drop these beliefs, because to do that is, in some sense, to call the belief into doubt. And what's at issue is whether they should call the belief into doubt. So what they need is some kind of disposition to replace a belief that p with an attitude that p is highly probable, and this disposition should correlate with the cases where taking p for granted will not maximise expected utility. This disposition will be a kind of vigilance. As Sperber and his collaborators show, we need some notion of vigilance to explain a lot of different aspects of epistemic evaluation. I think that notion can be usefully pressed into service here.[2]

If you need something like vigilance in your theory of belief, then you have to allow that vigilance might fail. Maybe some irrational

2 Kenneth Boyd (2016) suggests a somewhat similar role for vigilance in the course of defending an interest-invariant epistemic theory. Obviously I don't agree with his conclusions, but my use of Sperber's work does echo his.

dispositions can be traced to that failure, and not to any propositional attitude the decider has. For example, if Coraline systematically fails to be vigilant when exactly one billion dollars is at stake, then we might want to say that her belief in p is still rational, and she is practically, rather than theoretically, irrational. (Why could this happen? Perhaps she thinks of Dr Evil every time she hears the phrase "One billion dollars", and this distractor prevents her normally reliable skill of being vigilant from kicking in.)

If one tries to turn the vague talk of patterns of bets involving one proposition or another into a reductive analysis of when one particular belief is irrational, one will inevitably run into hard cases where a decider has multiple failures. We can't say that what makes Coraline's belief about the envelope, and not her belief about the battle, irrational is that if you replaced the envelope, she would invariably have a rational disposition. After all, she might have some other irrational belief about whatever we replace the envelope with. Or she might have some failure of practical reasoning, like a vigilance failure. Any kind of universal claim, like that it is only bets about the envelope that she gets wrong, won't do the job we need.

In "Knowledge, Bets and Interests", I tried to use the machinery of credences to make something like this point (Weatherson, 2012). The idea was that Coraline's belief in p was rational because her belief just was her high credence in p, and that credence was rational. I still think that's approximately right, but it can't be the full story. For one thing, beliefs and credences aren't as closely connected metaphysically as this suggests. To have a belief in p isn't just to have a high credence, it's to be disposed to let p play a certain role. (This will become important in the next two sections.) For another thing, it is hard to identify precisely what a credence is in the case of an irrational agent. The usual ways we identify credences, via betting dispositions or representation theorems, assume away all irrationality. But an irrational person might still have some rational beliefs.

Attempts to generalise accounts of credences so that they cover the irrational person will end up saying something like what I've said about patterns. What it is to have credence 0.6 in p isn't to have a set of preferences that satisfies all the presuppositions of such and such a representation theorem, which in turn maps one's preferences onto a probability function and a family of utility functions such that

$Pr(p) = 0.6$. That can't be right because some people have credence about 0.6 in p while not uniformly conforming to these constraints. But what makes them intuitive cases of credence roughly 0.6 in p is that generally they behave like the perfectly rational person with credence 0.6 in p, and most of the exceptions are explained by other features of their cognitive system other than their attitude to p.

In other words, we don't have a full theory of credences for irrational beings right now, and when we get one, it won't be much simpler than the theory in terms of patterns and explanations I've offered here. So it's best for now to just understand belief in terms of a pattern of dispositions, and say that the belief is rational just in case that pattern is rational. And that might mean that on some occasions p-related activity is irrational even though the pattern of p-related activity is a rational pattern. Any given action, like any thing whatsoever, can be classified in any number of ways. What matters here is what explains the irrationality of a particular irrational act, and that will be a matter of which patterns of irrational dispositions the actor has.

However we explain Coraline's belief, the upshot is that she has a rational, true belief that is not knowledge. This is a novel kind of Dharmottara case (or Gettier case for folks who prefer that nomenclature). It's not the exact kind of case that Dharmottara originally described. Coraline doesn't infer anything about the Battle of Agincourt from a false belief. But it's a mistake to think that the class of rational, true beliefs that are not knowledge form a natural kind. In general, negatively defined classes are disjunctive; there are ever so many ways to not have a property. An upshot of this discussion of Coraline is that there is one more kind of Dharmottara case than was previously recognised. But as, for example, Williamson (2013) and Jennifer Nagel (2013) have shown, we already knew that this is a very disjunctive class. So the fact that it doesn't look anything like Dharmottara's example shouldn't make us doubt it is a rational, true belief that is not knowledge.

8.2 Coin Puzzles

So rational belief is not identical to rationally having the dispositions that constitute belief. But nor is rational belief a matter of rational

high credence. In this section and the next I'll argue that even rational credence 1 does not suffice for rational belief. Then in the next section I'll run through some relatively familiar arguments that no threshold short of 1 could suffice for belief. If the argument of this section or the next is successful, those 'familiar arguments' will be unnecessary. But the two arguments I'm about to give are controversial even by the standards of a book arguing for IRT, so I'm including them as backups.

The point of these sections is primarily normative, but it should have metaphysical consequences. I'm interested in arguing against the 'Lockean' thesis that to believe p just is to have a high credence in p. Normally, this threshold of high enough belief for credence is taken to be interest-invariant, so this is a rival to IRT. But there is some variation in the literature about whether the phrase *the Lockean thesis* refers to a metaphysical claim, i.e., belief is high credence, or a normative claim, i.e., rational belief is rational high credence. Since everyone who accepts the metaphysical claim also accepts the normative claim, and usually takes it to be a consequence of the metaphysical claim, arguing against the normative claim is a way of arguing against the metaphysical claim. This section and the next argue that no matter how high the Lockean sets the threshold, their theory fails, since rational credence 1 does not entail rational belief. In Section 8.4, I'll go over puzzles that arise for Lockean theories that set the threshold below one.

The first puzzle for Lockeans comes from an argument that Williamson (2007) made about certain kinds of infinitary events. A fair coin is about to be tossed. It will be tossed repeatedly until it lands heads twice. The coin tosses will get faster and faster, so even if there is an infinite sequence of tosses, it will finish in a finite time. (This isn't physically realistic, but this need not detain us. All that will really matter for the example is that someone could believe this will happen, and it's physically possible that someone has that belief.)

Consider the following three propositions

A. At least one of the coin tosses will land either heads or tails.

B. At least one of the coin tosses will land heads.

C. At least one of the coin tosses after the first toss will land heads.

So if the first coin toss lands heads, and the rest land tails, B is true and C is false.

Now consider a few versions of the Red-Blue game (perhaps played by someone who takes this to be a realistic scenario). In the first instance, the red sentence says that B is true, and the blue sentence says that C is true. In the second instance, the red sentence says that A is true, and the blue sentence says that B is true. In both cases, it seems that the unique rational play is Red-True. But it's really hard to explain this in a way consistent with the Lockean view.

Williamson argues that we have good reason to believe that the probability of all three sentences is 1. For B to be false requires C to be false, and for one more coin toss to land tails. So the probability that B is false is one-half the probability that C is false. But we also have good reason to believe that the probabilities of B and C are the same. In both cases, they are false if a countable infinity of coin tosses land tails. Assuming that the probability of some sequence having a property supervenes on the probabilities of individual events in that sequence (conditional, perhaps, on other events in the sequence), it follows that the probabilities of B and C are identical. The only way for the probability that B is false to be half the probability that C is false, while B and C have the same probability, is for both of them to have probability 1. Since the probability of A is at least as high as the probability of B (since it is true whenever B is true, but not conversely), it follows that the probability of all three is 1.

Since betting on A weakly dominates betting on B, and betting on B weakly dominates betting on C, we shouldn't have the same attitudes towards bets on these three propositions. Given a choice between betting on B and betting on C, we should prefer to bet on B since there is no way that could make us worse off, and some way it could make us better off. Given that choice, we should prefer to bet on B (i.e., play Red-True when B and C are expressed by the red and blue sentences), because it might be that B is true and C false.

Assume (something the Lockean may not wish to acknowledge) that to say something might be the case is to reject believing its negation. Then a rational person faced with these choices will not believe *Either B is false or C is true*; they will take its negation to be possible. But that proposition is at least as probable as C, so it too has probability 1. So probability 1 does not suffice for belief. This is a real problem for the Lockean—no probability suffices for belief, not even probability 1.

8.3 Playing Games

Some people might be nervous about resting too much weight on infinitary examples like the coin sequence. So I'll show how the same puzzle arises in a simple, and finite, game.[3] 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 the underlying philosophical points are important, and it is easy to get confused about them.[4] So I'll set the basic points slowly, and then circle back to the puzzle for the Lockeans.[5]

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 Table 8.1.

Table 8.1 An underspecified decision problem.

	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.

3 This section is based on material from Weatherson (2016a: §1).

4 At least, I used to get these points all wrong, and that's got to be evidence they are easy to get confused about, right?

5 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.

Now compare that to a simple game. The players are Row and Column; Row will choose a row, Column will choose a column, and then the payouts will be given by the cell at the row and column's intersection. 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 Table 8.2. (As always, Row's payouts are given first.)

Table 8.2 A simple game.

	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.

The game in Table 8.2 is just a variant of the decision problem in Table 8.1. The relevant states of the world are choices of Column. Unlike the decision problem, there is a determinate answer to what Row should do in the game. More importantly for present purposes, the game can be solved without explicitly saying anything about probabilities. This is because we deduce all we need to know about probabilities from the assumption that Column is rational. Since Column is rational, they will play S_2. Since Column will play S_2, Row should play A_2.

Looking at games this way helps us understand why theorists sometimes think of game theory as "interactive epistemology" (Aumann, 1999). The theorist's work is to solve for what a rational agent should think other rational agents in the game should do. This is why game theory makes heavy use of equilibrium concepts. As theorists, we adopt a theory of rational choice, and see what happens if that theory is common ground amongst the players. In effect, we treat *rationality* as an unknown variable that we solve for given premises about which choices are rational in which games.[6] Not surprisingly, there are going to be multiple solutions to the puzzles we face.

6 If we're solving for a variable, what are the equations we're using as input. The standard methodology is to say they are intuitions. Game theorists make as much

This way of thinking naturally leads to the epistemological interpretation of mixed strategies. The most important solution concept in modern game theory is the Nash equilibrium. 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. Table 8.3 is one simple example.

Table 8.3 Death in Damascus as a game.

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

The only Nash equilibrium for this game is that Row plays a mixed strategy playing both A_1 and A_2 with probability ½, while Column plays the mixed strategy that gives S_1 probability 0.55, and S_2 with probability 0.45.

Now what is a mixed strategy? The *metaphysical* interpretation of mixed strategies is that players use some randomising device to pick what to do. This interpretation is often implicit in the way many textbooks introduce mixed strategies.

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

> 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: 57–58)

use of intuitions analytic philosophers. See, for example, Cho and Kreps (1987).

For our purposes, the important thing about the epistemological interpretation of mixed strategies is that it allows us to 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 game I'll call Up-Down.[7] Informally, in this game *A* and *B* must each play a card with an arrow pointing up, or a card with an arrow pointing down. I will capitalise *A*'s moves, i.e., *A* can play UP or DOWN, and italicise *B*'s moves, i.e., *B* can play *up* or *down*. If at least one player plays a card with an arrow facing up, each player gets $1. If two cards with arrows facing down 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, the payouts are given in Table 8.4, with *A* on the row and *B* on the column.

Table 8.4 The Up-Down game.

	up	*down*
UP	1, 1	1, 1
DOWN	1, 1	0, 0

I'll first work through Up-Down assuming Uniqueness: the epistemological theory that 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 align with those, such as Jill North (2010) and Miriam Schoenfield (2013), who reject this view. For now, I'll assume Uniqueness holds because it simplifies the analysis I'm about to offer; later, we'll relax the assumption.

Up-Down is symmetric. So given Uniqueness, *A* and *B* should have the same probability of playing UP/*up*. Call this common probability x. It cannot be that $x < 1$. A*'s expected return from UP is 1, while the expected return from DOWN is* x. *If* $x^* < 1$ and *A* is rational, they'll definitely play UP. If *A* will definitely play UP, the probability they'll play UP is 1, contradicting the assumption that $x < 1$.

7 In earlier work I'd called it Red-Green, but this is too easily confused with the Red-Blue game that plays such an important role in *Chapter 2*.

So we know $x = 1$. Arguably, we don't know that A will play UP. Assume we could know this. Whatever reason we would have for concluding that would be a reason for any rational person to conclude that B will play *up*. A is rational, so A will conclude this. So A's expected return from either strategy is 1. So A should be indifferent between UP and DOWN. Since all we know about A is that they are rational, and we know they are indifferent between UP and DOWN, we can't conclude, i.e., can't know, they will play UP.

There is an obvious objection to this argument. At one point I moved from the claim that A's expected return from UP and DOWN is the same, to the conclusion that A has just as much reason to play UP and DOWN. That looks like it is assuming that expected utility maximisation is the full theory of rationality. That, in turn, is something we might want to question.

In Chapter 6 I said that expected utility maximisation can't be the right theory of decision for agents who face non-trivial comptutational costs. This shouldn't be relevant here. A and B face pretty simple computations, and we can assume that the cost of those computations is negligible for each of them.

A more serious objection is that A has a reason beyond utility maximisation to play UP, namely that UP weakly dominates DOWN. After all, there's one possibility on the table where UP does better than DOWN, and none where DOWN does better. So perhaps even if UP and DOWN have the same expected utility, there is a reason to play UP.

As I've set up this game, this isn't actually an extra reason A has. To see this, it helps to compare the case to the kinds of games where Stalnaker (in the papers cited above) thinks that weak dominance does provide a distinct reason to make a choice. He is talking about games where the agents' attitude towards the possible payouts is different to their attitude towards each other. For example, the players may have common knowledge of the payouts, but only common belief in the rationality of each other. Or perhaps they even have rational, true belief in the rationality of each other, but crucially not knowledge. If that's right, but only if that's right, then it makes sense to use weak dominance reasoning.

The key motivation behind weak dominance reasoning is that taking a weakly dominated option is a needless risk. If UP will definitely return

1, while DOWN may return 0, then DOWN is risky in a way that UP is not. The notion of *risk* here need not be understood probabilistically. Even if it the probability that DOWN will return 1 is 1, there is still that payout of 0 sitting on the table, and so there is a risk.

Here we need to slow down. There is no outcome on the table where UP returns 1. But if the table is wrong, then UP might return 0. It might return anything at all. The only way that DOWN is risky while UP is not is if there is no risk that the table is mistaken.

Now one might object by pointing out that we stipulated A knows the table is correct and cannot be mistaken. We also stipulated that A knows that B is rational. So if rationality implies playing UP/*up*, there is no way that DOWN can return 0.

This is why Stalnaker's assumption that there is an asymmetry between the players' attitude towards the table and towards each other matters. If the players have a stronger attitude towards the rationality of each other than towards the correctness of the table, there is a sense in which irrational outcomes on the table are more of a risk than outcomes that are not on the table.

However, if the players think the players being irrational is exactly as live a possibility as the table being mistaken, then it is unreasonable to treat outcomes on the table which are only reached when the players are irrational as more relevant to decisions than outcomes not on the table at all.

That's why weak dominance reasoning is inappropriate in the Up-Down game. In some sense there is a risk DOWN could lead to a payout of 0. B might make an irrational move, even though, by stipulation, A knows that they will not. In the very same sense, there is a risk UP could lead to a payout of 0. The table could be wrong, even though A knows that it is not.

That's why the possibility of weak dominance reasoning doesn't undermine the reductio argument I've offered against UP/*up* being the uniquely rational play. It also helps us see why we ultimately don't need the assumption of Uniqueness to generate the objection.

Let's state the argument more carefully without Uniqueness. Assume, again for reductio, that some rational person C has credence $\varepsilon > 0$ that A will play DOWN. (It could be that C is a theorist, like us, or they could be one of the players.) We will now try to build a full model of C's attitudes towards the game.

Since it is common ground that A is an expected utility maximiser, C must have at least credence ε that A has credence 1 that B will play *up*. Is this coherent?

One reason to think not is that even without Uniqueness, it is strange to think that one rational agent could regard a possibility as infinitely less likely than another, given the exact same evidence.

Another reason to think this combination of views is incoherent is that without Uniqueness, the possibility of weak dominance reasoning comes back. If C has credence ε that A will play DOWN, then it is consistent with B's rationality that B has credence ε that A will play DOWN. Somehow C must have credence 1 that B does not have the same credences they do about what A will do, even though they and B have exactly the same evidence.

Uniqueness implies that C should have credence 1 that B will have the same credences as they do. I think Uniqueness is wrong, so I don't think that's a plausible constraint. But it's another thing to say that C should have credence 0 that someone in the same evidential situation as them has the same credences.

So even without Uniqueness, there are two reasons to think that it is wrong to have credence $\varepsilon > 0$ that A will play DOWN. Further, the argument that we can't know A will play UP did not rely on Uniqueness. So this is a case where 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 UP, like whether the coin will ever land tails, is a case where belief comes apart from high credence, even if by high credence we literally mean credence 1. This is a problem for the Lockean, and, like Williamson's coin, it is also a problem for the view that belief is credence 1.

8.4 Puzzles for Lockeans

I've already mentioned two classes of puzzles, those to do with infinite sequences of coin tosses and those to do with weak dominance in games. But there are other puzzles that apply especially to the kind of Lockean who identifies belief with credence above some non-maximal, interest-invariant, threshold.

8.4.1 Arbitrariness

The first problem for the Lockeans, and in a way the deepest, is that it makes the boundary between belief and non-belief arbitrary. This is a point that was well made some years ago now by Stalnaker (1984: 91). Unless these numbers are made salient by the environment, there is no special difference between believing p to degree 0.9876 and believing it to degree 0.9875. But if the belief threshold is 0.98755, this will be the difference between believing p and not believing it, which is an important difference.

The usual response to this is to say that the boundary is vague.[8] This won't help at all on theories of vagueness which endorse classical logic, like epistemicism (Williamson, 1994), or supervaluationism, or my preferred comparative truth theory (Weatherson, 2005b). In any of those theories there will still be a true existential claim that the threshold exists and is unimportant.

Even without settling what the right theory of vagueness is, we can see why this can't be right by thinking about what it means to say that a boundary is a vague point on a scale. Most comparative adjectives are vague, and the vagueness consists in which vague point on a scale is the boundary for their application. For example, whether a day is hot depends on whether it is above some vague point on a temperature scale. Vague comparative adjectives like 'hot' don't enter into non-trivial lawlike generalisations. There are laws involving the underlying scale, i.e., temperature, but no laws that are distinctively about the days that are hot. The most you can do is give some kind of generic claim. For instance, you can say that hot days are exhausting, or that electricity use is higher on hot days. But these are generics, and the interesting law-like claims will involve degrees of heat, not the hot/non-hot binary.

It's a fairly central presupposition of this book that belief is more connected to lawlike psychological generalisations than these mere generics. Folk psychology is full of lawlike generalisations that are essentially about belief. These are social science laws, not laws of fundamental physics, so the laws in question with be exception-ridden,

8 Versions of this response are made by Richard Foley (1993: Ch. 4), Daniel Hunter (1996), and Matthew Lee (2017b).

ceteris paribus laws. But they are laws nonetheless; they are explanatory and counterfactually resilient.

The Lockean fundamentally doesn't believe that these generalisations of folk psychology are anything more than generics, so this is a somewhat question-begging argument. The Lockean thinks the real laws are about credences, just like the real laws about hot days concern the underlying temperature scale. So my assumption that there are folk psychological laws about belief is strictly speaking question-begging. Nonetheless, it is true. I suspect any argument I could give for it would be less plausible than simply stating the claim, so I won't really try to argue for it. What I will do is illustrate why I believe it, and hopefully remind you why you believe it too.

Start by considering this generalisation.

- If someone wants an outcome O, and they believe that doing X is the only way to get O, and they believe that doing X will neither incur any costs that are large in comparison to how good O is, nor prevent them being able to do something that brings about some other outcome that is comparatively good, then they will do X.

This isn't a universal—some people are just practically irrational. But it's stronger than just a generic claim about high temperatures. It would still be true if the world were different in ever so many ways, and in cases where the person does X, this generalisation is part of the explanation for why they do X.

The Lockean denies almost all of that. They say this principle has widespread counterexamples, even among rational agents. Even when it is true, it isn't explanatory. Rather, it is a summary of some genuinely explanatory claims about the relationship between credence and action.

For example, the Lockean thinks that someone in Blaise's situation satisfies all the antecedents and qualifications in the principle. They want the child to have a moment of happiness. They believe (i.e., have a very high credence that) taking the bet will bring about this outcome, will have no costs at all, and will not prevent them doing anything else. Yet they will not think that people in Blaise's situation will generally take the bet, or that it would be rational for them to take the bet, or that taking the bet is explained by these high credences.

That's what's bad about making the belief/non-belief distinction arbitrary. It means that generalisations about belief are going to be not particularly explanatory, and are going to have systematic (and highly rational) exceptions. We should expect more out of a theory of belief.

8.4.2 Correctness

I've talked about this one a bit in Section 3.7.1, so I'll be brief here. Beliefs have correctness conditions. To believe p when p is false is to make a mistake. That might be an excusable mistake, or even a rational mistake, but it is a mistake. On the other hand, having an arbitrarily high credence in p when p turns out to be false is not a mistake. So having high credence in p is not the same as believing p.

Matthew Lee (2017a) argues that the versions of this argument by Jacob Ross and Mark Schroeder (2014) and Jeremy Fantl and Matthew McGrath (2009) are incomplete because they don't provide a conclusive case for the premise that having a high credence in a falsehood is not a mistake. But this gap can be plugged. Imagine a scientist, call her Marie, who knows the correct theory of chance for a given situation. She knows that the chance of p obtaining is 0.999. (If you think the belief/non-belief threshold is greater than 0.999, just increase this number, and change the resulting dialogue accordingly.) And her credence in p is 0.999, because her credences track what she knows about chances. She has the following exchange with an assistant.

> ASSISTANT: Will p happen?
> MARIE: Probably. It might not, but there is only a one in a thousand chance of that. So p will probably happen.

To their surprise, p does not happen. But Marie did not make any kind of mistake here. Indeed, her answer to the assistant's question was exactly right. But if the Lockean theory of belief is right, and false beliefs are mistakes, then Marie did make a mistake. So the Lockean theory of belief is not right.

8.4.3 Moorean Paradoxes

The Lockean says other strange things about Marie. By hypothesis, she believes that p will obtain. Yet she certainly seems sincere when she

says it might not happen. So she believes both p and it might not be that p. This looks like a Moore-paradoxical belief, yet in context it seems completely banal.

The same thing goes for Chamira. Does she believe the Battle of Agincourt was in 1415? Yes, say the Lockeans. Does she also believe that it might not have been in 1415? Yes, say the Lockeans, that is why it was rational of her to play Red-True, and it would have been irrational to play Blue-True. So she believes both that something is the case, and that it might not be the case. This seems irrational, but Lockeans insist that it is perfectly consistent with her being a model of rationality.

Back in Section 2.3.1 I argued that this kind of thing would be a problem for any kind of orthodox theory. And in some sense all I'm doing here is noting that the Lockean really is a kind of orthodox theorist. But the argument that the Lockean is committed to the rationality of Moore-paradoxical claims doesn't rely on those earlier arguments; it's a direct consequence of their view applied to simple cases like Marie and Chamira.

8.4.4 Closure and the Lockean Theory

The Lockean theory makes an implausible prediction about conjunction.[9] It says that someone can believe two conjuncts, yet actively refuse to believe the conjunction. Here is how Stalnaker puts the point.

> Reasoning in this way from accepted premises to their deductive consequences (p, also q, therefore r) does seem perfectly straightforward. Someone may object to one of the premises, or to the validity of the argument, but one could not intelligibly agree that the premises are each acceptable and the argument valid, while objecting to the acceptability of the conclusion. (Stalnaker, 1984: 92)

On the Lockean view, this happens all the time, and is intelligible. According to the Lockeans, it is easy to find triples $\langle S, A, B\rangle$ such that:

- S is a rational agent.
- A and B are propositions.
- S believes A and believes B.

9 This subsection draws on material from Weatherson (2016a).

- S does not believe $A \wedge B$.
- S knows that she has all these states, and consciously reflectively endorses them.

One argument against the Lockean is that there are no such triples, at least when S is rational. That's what I think. Even if I'm wrong, there is a separate argument against the Lockean. The Lockean doesn't just think these triples are possible, they think they are common. That's because for any $t \in (0, 1)$ you care to pick, triples of the form $\langle S, C, D \rangle$ are common.

- S is a rational agent.
- C and 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.

David Christensen (2005) argues from considerations about the preface paradox to the conclusion that triples like $\langle S, A, B \rangle$ are possible. His argument is non-constructive; he doesn't state a particular triple that clearly satisfies all the constraints, just argues that one must exist. I'm sceptical about that argument, but even if it worked, it wouldn't show what's needed. What's needed is that triples satisfying the constraints I set out for $\langle S, A, B \rangle$ are just as common as triples satisfying the constraints I set out for $\langle S, C, D \rangle$, for at least some value t. Considerations about esoteric cases like the preface paradox can't show that, and I haven't seen any other argument that even attempts to show it.

8.5 Solving the Challenges

Critiquing other theories for their inability to meet a challenge that one's own theory cannot meet is unfair. So I'll conclude this chapter by showing that the six problems I have presented for Lockeans do not

pose a problem for my interest-relative theory of (rational) belief. I've already discussed the points about correctness in Section 3.7.1, and about closure in Chapter 4 and Chapter 6, and there isn't much to be added. However, I would like to briefly touch upon the remaining four problems.

8.5.1 Coins

To believe p, one must have a disposition to take it for granted. A rational person prefers to bet on logically weaker propositions instead of logically stronger ones in the coin case. They would not take the logically stronger propositions for granted because if they did, they would be indifferent between the bets. Therefore, they would not believe that one of the coin tosses after the second will land heads or even that one of the coin tosses after the first will land heads. This is the correct outcome. The rational person assigns probability one to these propositions but does not believe them.

8.5.2 Games

In the Up-Down game, if the rational person believed that the other player would play *up*, they would be indifferent between UP and DOWN. But it's irrational to be indifferent between those options, so they wouldn't have the belief. They will think the probability that the other person will play UP/*up* is one—what else could it be? But they will not believe it on pain of incoherence.

8.5.3 Arbitrariness

According to IRT, the difference between belief and non-belief is the difference between willingness and unwillingness to take something as given in inquiry. This is far from an arbitrary difference. Moreover, it is a difference that supports lawlike generalisations. If someone believes that p, and believes that given p, A is better than B, they will prefer A to B. This isn't a universal truth; people make mistakes. But nor is it merely a statistical generalisation. Counterexamples to it are things to be explained, while instances are explained by the underlying pattern.

8.5.4 Moore

In many ways the guiding aim of this project was to avoid the kind of Moore-paradoxicality the Lockean falls into. So it shouldn't be a surprise that we avoid it here. If someone shouldn't do something because *p* might be false, that's conclusive evidence that they don't know that *p*. And it's conclusive evidence that either they don't rationally believe *p*, or they are making some very serious mistake in their reasoning. In the latter case, the reason they are making a mistake is not that *p* might be false, but that they have a seriously mistaken belief about the kind of choice they are facing. So we can never say that someone knows, or rationally believes, *p*, but their choice is irrational because *p* might be false.