Knowledge

A Human Interest Story

Brian Weatherson

https://www.openbookpublishers.com

©2024 Brian Weatherson

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

3. Belief

3.1 Beliefs and Interests

One core premise of this book is that someone knows something only if they properly take it to be settled. Taking something to be settled is what we might call believing it. Or, at least, it's a philosophically significant precisification of the notion of belief. Since belief and settling will play such an important role in the rest of this book, I'm going to discuss them here before we turn to knowledge.

The theory in this chapter owes a lot to proposals by Dorit Ganson (2008, 2019). Like her, I'm going to develop a theory where we first say what it is to have a belief in normal cases, then include an exception clause for what happens in special cases, such as high-stakes or long-odds cases. The details will differ in some respects, but the underlying architecture will be the same.

And it also owes a lot to work by Jonathan Weisberg (2013, 2020). Believing something is a matter of being willing to use that thing as an input to deliberation.[1] If we assume perfect rationality, it will often be possible to compute what inputs a thinker is using from the outputs of their deliberation. But it's a bad idea to assume perfect rationality in the general case, and without that assumption the inputs and outputs to deliberation can be arbitrarily far apart. And when they are, it's the inputs that matter to what someone believes. Here's how Julia Staffel puts the idea.

---

1   In earlier work I'd identified beliefs with something that we computed from the outputs of deliberation. This was a mistake; I should have been focussing on the inputs not the outputs. I'll say much more in Chapter 7 about how my views on this point have changed.

> One of the most important differences between outright beliefs and credences is how they behave in reasoning. If someone relies on an outright belief in p in reasoning, the person takes p for granted, or treats p as true. The possibility that $\neg p$ is ruled out. By contrast, if someone reasons with a high credence in p, they don't take p for granted. The possibility that p might be false is not ruled out. (Staffel, 2019: 939)

What's essential to belief is that to believe something is to be willing to use it as a starting point in deliberation. That slogan needs a lot of qualification to be a theory, but as a slogan it isn't a bad starting point.

Before we get too deep into this, I need to pause over a terminological point. When I talk about belief here, I mean to talk about the psychological aspect of knowledge. Roughly, that is, I'm talking about the mental state which is such that when things go well the thinker has knowledge, and which is indistinguishable from knowledge from the thinker's perspective. I'm not interested here in how closely this notion tracks the notion we pick out in ordinary language with words like 'believes' or 'thinks'.

This caveat is important because of a notable recent argument that belief is *weak* (Hawthorne, Rothschild, and Spectre, 2016). Imagine that some panellists on a TV show are discussing the upcoming Champions League season. They are asked who will win the League this year, and one of them says "I think Tranmere will win". And without theorising about this too much, assume this is an appropriate thing to say given their credal states and the situation they are in. Now see what happens to this case when we adopt two more premises. First, this is an honest and sincere self-report: they do, as we'd ordinarily say, think that Tranmere will win. Second, 'think' in English means *believes*. So this person believes Tranmere will win. Note though that in the circumstances of the TV show, they could say "I think Tranmere will win" even if they think Tranmere is merely the most likely team to win, which might happen even if they think the probability of that is very low. (If there are n teams in the Champions League, and who knows what value n will be when you're reading this, their credence that Tranmere will win could be maximal even if it is above 1 in n by an arbitrarily small amount.) Yet surely this person would not, at least responsibly, take Tranmere's winning to be a starting point in deliberation.

Now there are a lot of things we could say about that argument. I wouldn't want to sign up for either of the two premises that I mentioned in the middle of the paragraph. I'm sympathetic to the criticisms of the argument that Timothy Williamson makes in "Knowledge, Credence, and the Strength of Belief" (Williamson, forthcoming). For now, though, I just want to note that this is a discussion of a separate topic to the one I'm discussing. And in identifying the topic as I have, I'm working within a very standard, and very long, tradition. Here's Robert Pasnau, responding to a similar kind of challenge in the context of interpreting historical figures.

> I do not know of any historical figure who resists the idea that we can identify a kind of mental state, in the vicinity of assent, which can serve as a component in analyzing what it is to be in some more exalted epistemic state, in the vicinity of knowledge. What that component state gets called varies from century to century and from author to author. For Buridan, for instance, it will not be called *opinio*, because "*opinio* signifies a defect from scientia in some way" (Summulae VIII.4.4, trans. p. 710). But this is just a point about that Latin word, as it gets used at that moment in time, and goes no deeper than the analogous observation today that a *guess* cannot count as *knowledge*, no matter what gets added to it. Accordingly, throughout these lectures, I use 'belief' to pick out the mental state that is a constituent in the epistemically ideal state of *scientia* and so on, without fussing over whether 'belief' corresponds to *assensus*, *credere*, *opinio*, and so on. (Pasnau, 2017: 219)

I agree with all of that except possibly for the claim that belief is strictly speaking a constituent of *scientia*, or of knowledge. I want to leave open, at least at this stage, a knowledge-first account where belief is something like attempted knowledge. If that's right, knowledge would be a constituent of belief, and not vice versa. What's crucial is that there are close, even analytic, ties between belief as it's being used here and knowledge. Since our TV panellist can't know, and can't reasonably think they know, that Tranmere will win, their expression can't be an expression of belief, in this philosophically significant sense, that Tranmere will win.

Here's another way to put the point. It's a starting point in a lot of work in action theory that there is a true principle somewhere in the vicinity of the following idea.

> Zach intends to do some action, *A*. And he believes that to do *A*,
> he must do *B*. Zach bears an interesting and important normative
> relationship to *B*. It is an action that he believes to facilitate his intended
> end, and something is going wrong, if he intends *A*, believes *B* to be
> necessary for *A*, has reflected clear-headedly on this fact, and yet still
> fails to intend to do *B*. (Schroeder, 2009: 223)

There are challenges about how to make this principle quite right in cases where Zach shouldn't intend to do *A*. If the 'belief is weak' thesis is correct, however, the whole tradition in action theory that Schroeder is here joining is fundamentally mistaken. From the intention to do *A*, and the best guess that the only way to do *A* is *B*, it does not follow at all that coherence requires intending to do *B*. Since I don't think that the entire literature on means-end coherence was based on fundamentally misunderstanding the nature of belief, I'm going to assume that we have a strong notion of belief. Just how it relates to the English words 'guess', 'think', and even 'believes' is left as an issue for another day.

3.2 Maps and Legends

Beliefs, Frank Ramsey famously said, are maps by which we steer (Ramsey, 1990: 146). This can be turned into an argument that belief should be interest-relative as well. This argument isn't quite right (contrary to my earlier views), but it's instructive to see why it goes wrong. First let's explore Ramsey's analogy a bit more closely.[2]

When I was growing up in car-dependent, suburban Melbourne, the main street directory that was used was the *Melways*. This was a several-hundred-page-thick book that most people kept a copy of in their car. It largely consisted of page after page of 1:20,000 scale maps of the Melbourne suburbs, plus more detailed maps of the inner city, and then progressively less detailed maps of the rural areas around Melbourne, the rest of Victoria, and finally of the rest of the country. And it was everywhere. It was common for store advertisements, party

---

2   The picture I'm sketching about the map-like nature of belief is similar to the one
     that Seth Yalcin has defended in two articles (2018 and, especially, 2021). That's
     not to say he would endorse any of the conclusions here, but simply to note that he
     has set out the the idea that belief is less like a map and more like an atlas, and put
     that idea to philosophical work.

invitations and event announcements to include the *Melways* page and grid coordinates of the location. In fact I was a little shocked when I moved to America and I found it was socially expected (in those pre-Google Maps days) that you would give people something like turn-by-turn directions to a location. I was used to just telling people where something was, i.e., giving them the *Melways* grid coordinates, and letting them use the map to get themselves there. The *Melways* really was, collectively, the map by which we steered.

But you wouldn't want to use it for everything. You wouldn't want to use it as a hiking map, for example. For one thing, it was much too heavy. For another, it was patchy on which walking trails it even included, and had almost no usable topographical information. You steer yourself by one map when you drive, and another map (or set of maps) when you hike. What one steers by should be a function of one's interests. And the same is true of belief. For most people, beliefs are interest-relative because to believe something is to steer yourself by a map that represents the world as being that way, and which map one will steer by is sensitive to one's interests.

Maybe you think this argument leans too heavily on Ramsey's analogy of beliefs and maps. But once you see the structure of the case, you can get more purely cognitive examples. (And this in turn helps us see the brilliance of Ramsey's metaphor.) If you or I were in Anisa's position, then we would not include the fact that the Battle of Agincourt was in 1415 on the map by which we steer through the Red-Blue game, even if we would typically include it on our map. When I'm reading the morning papers and thinking about the effects of some economic policy, such as a proposed minimum price for alcohol, I'll steer myself by the maps given in introductory economics texts. That is, I'll just use simple supply-demand graphs to predict the effects of the policy. Still, I won't always do that. For example, I won't do it when thinking about changes in the minimum wage, because systematic changes like that push simple models beyond their breaking point.[3] Or we can mix and

---

3   I've said in the text that I believe that simple supply-demand models are right for some purposes. At least, I implied that when I said I steer by them, and that beliefs are maps by which we steer. Some philosophers think this is wrong, and that one only ever accepts these simple models, rather than believes them. Once we allow beliefs to be interest-relative, this role for the belief/acceptance distinction seems

match the practical and the theoretical. If there is a proposed price floor on something widely traded (like electricity), and my predictions about the effects of this change have even a small practical significance (e.g., I'm thinking about whether my small business should lock in the price it purchases electricity at for three years), then I might not use the simple model. In this case the combination of theoretical and practical interests will change which map I steer by, i.e., what I believe, even if neither interest on their own would have been enough to bring about a change.

So it looks like belief is interest-relative, and that's for deep reasons about the role that belief plays in our cognitive economy. To believe something is to steer by a map that represents it as true. To steer by it, in this sense, is to take it as given in our inquiries. For normal people, what is taken as given is dependent on what question one is interested in. So for normal people, belief is interest-relative. I used to think that this could be extended to an argument that it was part of the metaphysics of belief that it was interest-relative. But as we'll see in the next section, that isn't quite right. The restriction to 'normal people' a couple of sentences back turns out to be essential, and this creates complications.

3.3 Belief and Stubbornness

Things get complicated when we stop focussing on what normal (or normal-ish) people do, and think about less common reactions. So consider a person, call them Stubbie, who uses the same maps and models for every task. He uses the *Melways* for hiking, he makes macro-economic forecasts using simple supply-demand models, he takes ordinary knowledge for granted in high-stakes and long-odds cases, and so on. And he does this even though he knows full well that there are excellent reasons to be more flexible. What should we say about Stubbie?

I think we should say that Stubbie is irrationally stubborn, and part of his irrationality consists in steering by the same map, in holding onto the same beliefs, in situations where this is uncalled for. Stubbie acts as if simple supply-demand models are predictive in complicated

---

to go away. A lot of what are commonly called acceptances are, in my theory, beliefs that are highly sensitive to changes in interests.

situations, and as if the *Melways* has all the information a hiker needs. Neither of these things are true, and Stubbie should know they aren't, but our theory of belief had better allow for some irrational practices that could only be rationalised by false assumptions.

Stubbie's example shows that while one's beliefs should be interest-relative, they need not be. One should steer by a map suitable to the circumstances. If one stubbornly steers by the same map come what may, the fact that it would be advisable to steer by different maps at different times does not affect what one believes. Stubbie really is steering himself by the *Melways* when hiking, and he really believes the simple economic model he uses.

This shows that one can be a believer, without having those beliefs be sensitive to one's interests. That suggests that the interest-relativity of belief comes from the norms—how one should believe—not the metaphysics—what belief itself is.

There is another complicated variant of this example that raises deeper questions about the relationship between belief and interests. Imagine that Stubbie is disposed to keep taking what history books say about Agincourt for granted. Now he is faced with a decision where a lot rides on this practice. Perhaps he is playing a version of the Red-Blue game where the prize is $50,000, not $50. And the shock of having that much at stake causes him to reconsider. So he goes back to thinking it merely probable that the Battle of Agincourt was in 1415. This is not a case of interest-relativity of belief. Rather, it is like the kind of case Jennifer Nagel (2010) discusses, when she talks about beliefs being causally sensitive to interests. And this shows we have to be careful in order to be sure that a case of interest-sensitivity is really a case where belief is constitutively, and not merely causally, sensitive to interests.[4]

This version of Stubbie's case opens up the possibility that no beliefs are really interest-relative. Sometimes a change in circumstances might cause someone to change the map they steer by, but that's the only way that interests matter. I don't think this is right, but I'm much less confident of this than I am of most of the other claims in this book.

There are three significant differences between the way that interests change the beliefs of normal people to how they change Stubbie's beliefs.

---

4   In earlier work I was not careful on exactly this point. I'll say more about this in Chapter 7.

First, they are reversible. Someone who switches to a more complicated model, or to thinking that a source provides probability rather than knowledge, can easily switch back. Second, they are predictable. For a reasonably well-functioning thinker, we can say when they will switch maps. It will be when the stakes are high, or the odds are long, or the question pushes on the limitations of their models. Third, they are not emotionally loaded. The natural way to tell this variant of Stubbie's story involves shock; he feels the change in his attitude. But when you or I play the Red-Blue game, we switch from thinking something is true to thinking it is probable without any significant phenomenology. I think these three differences are enough to justify saying that in the normal case, the change of interests constitutes a change of beliefs, while in Stubbie's case, the change of interests merely causes a change of beliefs. And if that's right, the belief itself is interest-relative, in normal cases.

But whether we accept the argument of the last paragraph or not, it won't affect what we say about Anisa. She believes the Battle of Agincourt was in 1415. This belief is irrational; she should have switched to thinking it is merely probable that the battle was in 1415. The change in the rational status of her belief is constituted by, and not merely caused by, her change in interests. So interests can be constitutively relevant to rational belief, even when they don't affect belief.

The next two sections aim to turn these Ramseyan observations about the relationship between beliefs and interests into a theory of belief.

3.4 Taking as Given

To start towards a positive theory of belief, it helps to think about the following example, featuring a guy I'll call Sully. (This example is going to resemble the examples involving Renzo in Ross and Schroeder (2014), and at least for a while, my conclusions are going to resemble theirs as well.) Sully is a fan of the Boston Red Sox, and one of the happiest days of his life was when the Red Sox broke their 86-year-long curse, and won baseball's World Series in 2004. He knows, and hence believes, that the Red Sox won the World Series in 2004. He likes their chances to win again this year, because in Sully's heart, hope always springs eternal.

It's now the start of a new baseball season, and Sully is offered, for free, a choice between the following two bets.

- Bet A wins $50 if the Red Sox win the World Series this year, and nothing otherwise.

- Bet B wins $60 if the same team wins the World Series this year as won in 2004, and nothing otherwise.

For Sully, this choice is a no-brainer. If the Red Sox win this year, he wins more money taking B than A. If the Red Sox don't win this year, he gets nothing either way. So it's better to take B than A, and that's what he does.

What Sully has done here is use dominance reasoning, in particular weak dominance reasoning. One option weakly dominates another if it might have a better return, and can't have a worse return. Weak dominance is used as an analytical tool in game theory. It is also a form of inference that non-theorists, like Sully, can use. (Though unless they've taken a game theory course they might not use this phrase to describe it.)

Sully's case can be distinguished from that of his more anxious friend Mack. Mack is also a big Red Sox fan, and also looks back on that curse-busting World Series win with fondness. But if you offer Mack the choice between these two bets, he'll hesitate a bit. He'll wonder if he's really sure it was 2004 that the Red Sox won. Maybe it was 2005, he thinks. He'll eventually think that even if he's not completely sure that it was 2004, it was very likely 2004, and so it is very likely that bet B will do better, and that's what he will take.

Even if Sully and Mack end up at the same point, they have used very different forms of reasoning. Sully uses weak dominance reasoning, while Mack uses probabilistic reasoning. Sully takes the fact that the Red Sox won in 2004 as given, while Mack just takes it to be very likely. The big thing I want to rely on here is that these are very different psychological processes. Neither of these guys is doing something that approximates, or simplifies, the other; they both take bet B, but they get to that conclusion via very different routes.

There is a theoretical analogue to this psychological point. Many game theorists, perhaps most, think that weak dominance reasoning can be iterated more or less indefinitely. (That's not to say that they are right; I'm trying to make a point about conceptual distinctiveness here, not game theory.) But few if any think that likelihood reasoning can be

iterated indefinitely. This reflects the fact that they are very different kinds of reasoning. Dominance reasoning is pre-probabilistic.

Sully's reasoning isn't just dominance reasoning. It's dominance reasoning that relies on a contingent assumption, namely that the Red Sox won the World Series in 2004. When Sully reasons that A can't do better than B, he's not drawing any kind of logical or metaphysical point. It's logically and metaphysically possible that the Red Sox lost in 2004. For that matter, and this is a point Ganson (2019) stresses, it's logically and metaphysically possible that the payouts for A and B are other than what Sully thinks they are.

And though he might not make it explicit, at some level Sully surely knows this. If pushed, he'd endorse the conditional "If I've misremembered when the curse-busting World Series win was, and the Red Sox didn't win in 2004, then bet A might do better than bet B". So while he is disposed to use dominance reasoning in deciding whether to take A or B, this disposition rests on taking some facts about the world for granted.

Recall the disjunctive way that Sully reasoned. Either the Red Sox will win this year or they won't. Either way, I won't do better taking bet A, but I might do better taking bet B. So I'll take bet B. This reasoning—not just the reasons Sully has but his reasoning—can be appropriately represented by the kind of decision table that is familiar from decision theory or game theory.

Table 3.1 Betting on the Red Sox.

	Red Sox Win	Red Sox Don't Win
Take Bet A	$50	$0
Take Bet B	$60	$0

Focus for now on the columns in this table. Sully takes two possibilities seriously: that the Red Sox win this year, and that they don't. The 'possibilities' here are possibilities in the sense described by Humberstone (1981). They have content—in one of them the Red Sox win, in the other they don't, but they don't settle all facts. In the right-hand column, there is no fact of the matter about which other team wins the World Series. In neither column is there a fact of the matter about what Sully will have

for lunch tomorrow. If you want to think of these in terms of worlds, they are both very large sets of worlds, and within those sets there is a lot of variability.[5]

But there is more to the content of each column than what is explicitly represented in the header row. In each column, for example, the Red Sox won in 2004. That's why Sully can put those monetary payoffs into the cells. And in each column, the terms of the bet are as Sully knows that they are. In sets of worlds terms, the sets that are represented by the columns are exclusive, but far from exhaustive.

Consider those propositions which are true according to all of the columns in this table. Say a proposition is *taken as given* in a decision problem when it the decider treats one option as dominating another, and does so in virtue of a table in which that proposition is true in every column. Then here is one principle about belief that seems to be very plausible.

Given
S believes that *p* only if there is some possible decision problem such that S is disposed to take *p* as given when faced with that problem.

Given is logically weak in one respect, and strong in another. It only requires that S be willing to take *p* for granted in one possible choice. It doesn't have to be a likely, or even particularly realistic choice. Sully is unlikely to have strangers offer him these free money bets. Given how representationally sparse decision tables are, for something to be true in all columns of a decision table is a very strong claim. It doesn't suffice, for instance, for *p* to be true in some columns and false in none. Each column has to take a stance on *p*, and endorse it.

I will have much more to say about the relationship between decision tables like Table 4.1. First, however, I need to say more about belief. I used to think that Given, or something like it, could be strengthened into a biconditional, and from there we could get something like a functionalist analysis of belief. That turns out not quite to be right. Being disposed to sometimes take *p* as given is not sufficient for belief. If Anisa had played the Red-Blue game rationally, she would have lost any belief

---

5   Analysing these possibilities as sets of worlds is unhelpful when we want to use a model like this to represent modal or logical uncertainty. Still, it's often a helpful heuristic, and there isn't anything wrong with using a model that breaks down when applied outside its appropriate zone.

about when the Battle of Agincourt was. To explain cases like that, we need to expand our theory of belief.

3.5 Blocking Belief

Imagine a person, call him Erwin, who is made the offer Blaise is made, but declines it. He declines on the very sensible grounds that the Battle of Agincourt might not have been in 1415, and he does not want to run the risk of sending everyone to The Bad Place. If we stop our theory of belief with Given, then we have to say that Erwin has some kind of weird pragmatic incoherence. He believes that p, and wants what is best for everyone, but won't do the thing that will, given his beliefs, produce what is best for everyone. Declining the bet is not practically incoherent in this way. So Erwin does not believe that the Battle of Agincourt was in 1415. At least, he doesn't believe that at the time he is declining the bet.

So a theory of belief with any hope of being complete needs some supplementation. The idea I'll use is one that seems *prima facie* like it might apply without restriction. A little reflection, however, shows that it will ultimately need to be restricted, and the most natural restrictions are pragmatic.

Imagine that we don't ask Erwin whether he is prepared to bet the welfare of all of humanity on historical claims, but instead ask him a simple factual question H.

> H. How many (full) centuries has it been since the Battle of Agincourt?

Erwin will think to himself, "Well, the Battle of Agincourt was in 1415, and that's a bit over six hundred years ago, so that's six centuries. The answer is six." Now compare what happens if we ask him this slightly more convoluted question.

> I. If the Battle of Agincourt was in 1415, how many (full) centuries has it been since the Battle of Agincourt?

Erwin will give the same answer, i.e., six. And he will give it for basically the same reasons. Indeed, apart from the date of the Battle being one of his reasons in answering H, and not needed to answer I, he has the same reasons for answering the two questions with six. I mean that both in the sense that what justifies giving the answer six is the same for the

two questions, and in the sense that what causes him to answer six is the same for the two questions. (With the exception that the date of the battle is a reason in answering H, but not in answering I.)

Say that a person answers the questions *Q?* and *If p, Q?* in the same way if they offer the same answer to the two questions, and their reasons (in both senses) for these answers are the same except only that *p* is one of the reasons for their answer to *Q?*. Then here is a plausible principle about belief—albeit one that isn't going to be quite right.

> **Unrestricted Conditional Questions**
> If S believes that *p*, then for any question *Q?*, S is disposed to answer the questions *Q?* and *If p, Q?* the same way.

Note that in saying these questions are answered the same way, I really don't just mean that they get the same answers. I will offer the same answer to the questions *What is one plus one?* and *What is the largest n such that xn + yn = zn has positive integer solutions?*, but I don't answer these questions the same way. My reasons for the first answer are quite closely related to the fact that one plus one does equal two. My reasons for the second answer are almost wholly testimonial. So in the sense relevant to Unrestricted Conditional Questions, I do not answer each question the same way.

I'm understanding what a conditional question is in a particular way, one I'll describe in the next paragraph. I think this is how conditional questions usually work in English, so the shorthand *If p, Q?* that I'm using is not misleading. But I don't intend to defend a particular claim about the way natural language conditionals work. That would be another whole book (or more.) So I intend to use this shorthand *If p, Q?* somewhat stipulatively, as follows.

If p, Q? is the question *Q?* asked under the assumption that *p* can be taken as given. So the question *If p, how probable is q?* is asking for the conditional probability of *q* given *p*. The question *If p, which option is most useful?* is asking for a comparison of the conditional utilities of the various options. And the question *If p, must it be that q?* gets an affirmative answer if all the (salient) possibilities where *p* is true are ones where *q* is true. (So it becomes very close to asking if the material implication $p \supset q$ must be true.) Now it is notoriously difficult to connect these conditional questions with questions about the truth of

any conditional.[6] But I'm setting all those issues aside here. Everything that I say about conditional questions I could say, more verbosely, by making it explicit that they are to be understood as questions about conditional probability, conditional utility, conditional modality, and so on.

Now thinking about a few simple cases might make it seem that Unrestricted Conditional Questions is true. After all, there is something very odd about a counterexample to it. It would have to be a case where S believes that p, and there is a way they are disposed to get answer *If p, Q?*, i.e., to get from p to an answer to $Q?$, but they are not disposed to use that to answer $Q?$. That seems at best rather odd.

There is one potential counterexample that I don't think ultimately undermines Unrestricted Conditional Questions. There could be a case where I believe p, and p is relevant to $Q?$, but I don't realise its relevance. On the other hand, when I am explicitly asked *If p, Q?*, being reminded of p makes me see the connection, so I follow the natural path from p to an answer to $Q?$. These kinds of one-off performance errors are, sadly, easy to make. As long as they are one-off, they don't threaten the principle connecting dispositions.

A bigger problem comes from the two cases that I started the book with. If the Battle of Agincourt was in 1415, then Anisa maximises expected utility by playing Blue-True, and Blaise maximises expected utility by taking the bet. So answers to the conditional questions *If the Battle of Agincourt was in 1415, what options of Anisa's maximise expected utility?* and *If the Battle of Agincourt was in 1415, what option of Blaise's maximises expected utility?* are different to the answers to the corresponding unconditional questions. Or at least so say I, and hope you do too. So if Unrestricted Conditional Questions is true, then none of us have ever believed that the Battle of Agincourt was in 1415. That can't be right, so there must be some restriction on the principle.

Happily, a restriction isn't too hard to find. The principle just needs to be restricted to questions that the subject is currently taking an interest in. When we're thinking about questions like H and I, then we do have

---

6   See Lewis (1976, 1986) on the issues about conditional 'how probable' questions; Lewis (1988, 1996) on the issues about conditional 'how useful' questions; and Gillies (2010) on issues about modals in the consequent of conditional questions.

beliefs about when the Battle of Agincourt was. Were we to be placed in Anisa or Blaise's situation, or arguably when we even think about their situation, we lose this belief. So I suggest the following principle is true, and explains a lot of the cases that have been discussed so far.

> **Relevant Conditional Questions**
> If S believes that *p*, then for any question *Q?* that S is currently taking an interest in, S is disposed to answer the questions *Q?* and *If p, Q?* the same way.

As I argued in Section 2.5, whether one is interested in a question isn't just a matter of one's practical situation. One can be interested in a question because one is thinking about what to do should it arise, or because one is just naturally inquisitive. Many of the questions we're interested in are practical questions, but not all of them are.

I've argued that Given and Relevant Conditional Questions are necessary conditions on belief. Very roughly, I think they are jointly sufficient for belief. I say 'roughly' because I don't mean to take a stance on, say, whether animals have beliefs, or whether one can have singular thoughts about things one is not acquainted with. A more accurate claim is that if it is plausible that S is the kind of thing that can have beliefs, and *p* is the kind of thing it could in principle have beliefs about, and both Given and Relevant Conditional Questions are satisfied, then S believes that *p*.

Obviously neither Given nor Relevant Conditional Questions would be particularly helpful principles to use in providing a reductive physicalist account of mental content. They say something about necessary conditions for belief, but the statement of those conditions makes a lot of assumptions about other content-bearing states of the agent. So even if these conditions are individually necessary and jointly sufficient for belief, they wouldn't be any kind of analysis or reduction of belief.[7] But they could be part of a theory of belief, and the theory they are part of is helpful for seeing how beliefs and interests fit together.

---

7   Compare: One can consistently deny that any analysis or reduction of *knowledge* is possible and say that the condition *p is part of S's evidence* is both necessary and sufficient for S to know that *p*.

3.6 Questions and Conditional Questions

In the previous section I defended this principle:

> **Relevant Conditional Questions**
> If *S* believes that *p*, then for any question *Q?* that *S* is currently taking an interest in, *S* is disposed to answer the questions *Q?* and *If p, Q?* the same way.

To spell out what that principle amounts to, I need to say something about what questions are, and what conditional questions are. I'm going to say just enough about questions to understand the principle. This won't be anything like a full theory of questions. While much of what I say will draw on insights from theorists who have worked on questions in natural language, I'm not primarily interested in how questions are expressed in natural language. Rather, I'm interested in the contents of these questions. These contents are interesting because they can be the contents of mental states. For example, a cat can wonder where a mouse is hiding. There are deep and fascinating issues about how we can and do talk about the cat, and the cat's attitudes, but I'm more interested in the cat's relationship to the question *Where is the mouse hiding?* than I am in our talk about the cat.[8]

The simplest questions are true/false questions, like *Did the Boston Red Sox win the 2018 World Series?* These won't play a huge role in what follows, but they are important to have on the table. I am going to assume that whenever someone considers a proposition, and they don't take its truth value to be settled, they are interested in the question of whether it is true.

Next, there are quantitative questions, where the answer is some number or sequence of numbers.[9] One tricky thing about quantitative questions is that they may admit of imprecise answers, but need not. If

---

8　A useful introduction to ways in which questions are relevant to philosophy of language is the Stanford Encyclopedia article by Cross and Roelofsen (2018). A canonical text on the role of questions is Roberts (2012). Roberts originally circulated that paper in 1996. Since then it has influenced a huge range of works, including this one.

9　I'm including here any question that could be answered with a number or sequence of numbers, even if that would not be the most usual, or the most helpful, way to answer them. So *Where is Fenway Park?* is a quantitative question, because *42.3467° N, 71.097° W* is an answer, even if *The corner of Jersey St and Van Ness St* is a better answer.

I ask, "When does tonight's Red Sox game start?", an answer of "Seven" would usually be acceptable, even if the game actually starts at a few minutes after seven. That's because, I take it, the truth conditional content of the utterance "Seven" in this context is that tonight's Red Sox game starts at approximately seven, and I'm asking a question that admits of an approximate answer. I could have been asking a question where the only acceptable answer would be the time that the Red Sox game starts to the nearest minute, or even to the nearest second. And I could even have asked that question using those exact same words. (Though if I intended to ask the question about seconds, using these words would be extremely unlikely to result in communicative success.)

The main thing that matters for the purposes of this book is that the questions with different appropriate answers are different questions. Even if one would normally use the same words in English to express the questions, the fact that they have different acceptable answers shows that they are different questions. And as noted above, what really matters for this book is the mental representation of the contents of questions. There could be two people who we could report as wondering when tonight's Red Sox game starts, but one of them will cease wondering if they find out that it starts around seven, and the other still wonders which minute near seven it will start at. These people are wondering about different questions.

The more precise a numerical question one is considering, the fewer things one can rationally take for granted in trying to answer it. So the version of IRT I defend implies that the more precise a numerical question one is considering, the fewer things one knows. Or, to put the same point another way, the less precise a numerical question one is considering, the less impact interest-relativity has on knowledge. This will matter when thinking about how the theory applies to various examples. If we ascribe to a thinker an interest in an unrealistically precise question, we might draw implausible conclusions about what IRT says about them. But this isn't a consequence of IRT; it's a consequence of not getting clear about which question a thinker is considering.

Next, there are questions that ask to identify an individual or a class of individuals. A striking thing about these questions is that they often have so-called 'mention-some' readings. To understand what this means, compare these two little exchanges.

1. a. Who was in the Beatles?
    b. John Lennon was in the Beatles.

2. a. Where can I get good coffee in Melbourne?
    b. You can get good coffee at Market Lane.

There is something wrong with 1b as an answer to 1a. It's true that John Lennon was in the Beatles. But an ordinary use of 1a will be to ask for the names of everyone in the Beatles, not just one person in them. (There are exceptions, and it's a fascinating task for another day to work out when they occur.) On the other hand 2b is a perfectly good answer to 2a. (Or so I think, but my knowledge of Melbourne coffee is a little out of date.) It is definitely not necessary to properly answer 2a that one list every place in Melbourne where one can get good coffee. That could take some time. Moreover, 2b does not (on its most natural reading) imply that Market Lane is the only place in Melbourne to get good coffee.

An answer is a 'mention-some' answer when it does not imply exhaustivity in this sense. And a question admits of mention-some answers when it is properly answered with a mention-some answer. Lots of questions asking for individuals will be mention-some questions in this sense, but not all of them will. And, again, it is important to understand what kind of question is being asked to think about whether it is satisfactorily answered by an answer that does not imply completeness or exhaustiveness.

Next, there are questions with infinitivals, such as the following.

- When to visit Venice?

- How to climb Ben Nevis?

- What to do?

In most dialects of English, it is rare to use these to simply ask questions.[10] But they can be the complements of any number of verbs. Any of the three questions above, like any number of other questions with infinitivals, can complete sentences like:

- A doesn't know …

- B is wondering …

_____

10   My hunch is that there is quite a bit of dialectical variation here; I would need to do much more empirical research to back this up.

- C wants D to tell him ...

Mixing and matching the sentence fragments from the last two lists produces nine different sentences. Some examples of these are:

- C wants D to tell him how to climb Ben Nevis.

- A doesn't know what to do.

- B is wondering whether to visit Venice.

The philosophical work on these kinds of sentences has been almost exclusively focussed on just one of the nine sentences I just described: the one combining a knowledge verb with a 'how to' question. I suspect this is a mistake; what to say about 'know how' reports is going to have a lot in common with what to say about 'wondering when' reports. (Here I'm agreeing with Stanley (2011), though I'm about to disagree with him on a related point.)

There is a puzzle about why, in English, we cannot use these questions to complete sentences like:

- E believes ...

- F suspects ...

- G wants H to guess ...

I'm going to set that puzzle aside, as interesting as it is, and just focus on the sentences we can produce in English.

I'm going to call these questions with infinitivals *practical questions*. One thing to note about them is that they are usually mention-some. When I am wondering what to buy in the supermarket, and I resolve this by choosing one particular carton of eggs, I don't thereby imply that there is anything defective about the other cartons. I just choose some eggs.

For related reasons, answering a practical question like this is distinct from answering any question, or questions, about the modal status of different actions. Imagine that in the grip of choice-phobia I am stuck staring at the cartons of eggs, unable to decide which one to buy because they are all just alike. In that situation I might know that there is no carton such that it is what I should buy, and also that there are many cartons such that I could (rationally, morally) buy any one of them. But

there are so many, and they are so alike and I can't decide, so I don't know what to buy.[11]

Resolving this indecision will not involve accepting any modal proposition like *I should buy this carton in particular*. It had better not, because I really have no reason to accept any such proposition. Rather, it involves accepting a proposition like *I will buy this carton in particular*. I can accept that by simply buying the eggs. There were many other answers I could equally well have accepted, since there were many other cartons I could buy.[12]

Practical questions are distinct from questions about modals or utilities, but there will usually be a correlation between their answers. Usually, if someone asks you when to visit Venice, and there is one time in particular such that visiting then maximises expected utility, that's what you should tell them. That's when they should visit, and that's what to say when they ask you when to visit. Relatedly, practical questions can come in conditional form. We can utter sentences like the following in English.

- J asks K what to do if his patient has hepatitis.

And there is one feature of these sentences that needs noting. I don't know what to do if one's patient has hepatitis, so let's just say that K tells J to do X. What that means is not that in any situation where the patient has hepatitis, do X. If the patient's symptoms are confusing, it might be best to run more tests before doing X. What it does mean is that if the fact that the patient has hepatitis is taken as given, then do X. As always,

---

11   This discussion will probably remind many readers of the story of Buridan's ass, who was stuck between two equally appetising bales of hay. As Peter Adamson (2019: 453ff) points out, the connection of this example to Buridan is not the one philosophers usually assume. That is, it's not Buridan's example. An example of roughly this kind was earlier given by al-Ghazālī. And the example involving the ass was not given by Buridan at all, but by his opponents, objecting to Buridan's own equation of choice with judgment that something is best to do. That's the role the example will play a few times in this book, as a critique of theories that equate choice with formation of a belief about goodness. My earlier versions of IRT, which equated that choosing to do something with judging it has highest expected utility, will be among the theories thus targeted.

12   I'm here mildly disagreeing with Stanley (2011: Ch. 5) when he says that these questions with infinitival complements can be paraphrased using modals like 'should'. If 'will' just is the modal that gets used in the paraphrase, as Bhatt (1999) suggests, the spirit of Stanley's view is preserved, even if the letter isn't.

conditional questions should be understood as questions about what happens in scenarios where the condition in question is taken as given. And the constraint expressed by Relevant Conditional Questions is that whatever is known can be taken as given in just this sense.

3.7 A Million Dead End Streets

As I've noted already, the view I'm defending here is somewhat different from my earlier view. And it's helpful to understand the view of this book to lay out, in one place, the ways in which time has changed my views. Here is a somewhat simplified version of the view from "Can We Do without Pragmatic Encroachment?". Assume that S is interested in some quantitative questions and some alethic (i.e., yes/no) questions. Then the view was that S believes that *p* if and only if these two conditions are met.

1. For any quantitative question *Q?* that S is interested in, and any alethic question *A* that S is interested in, S's answers to the question *If A, Q?* and *If A and p, Q?* are the same.

2. S's credence in *p* is greater than 0.5.

It was assumed that S is always 'interested' in the null question *Is a tautology true?*, so one special instance of this is that S answers *Q?* and *If p, Q?* the same way. And it was assumed that S is an expected utility maximiser, so the practical question of what to do becomes just the quantitative question *Which of these options has the highest expected utility?*. There are bells and whistles, especially in thinking about the level of precision that goes along with the quantitative questions that S is interested in. (Draw these too fine, and S doesn't have beliefs, so you have to be a little careful here.) Even without those complications, I've said enough that you can see the basic view, and perhaps see its problems.

The biggest change from that view to the one I'm defending here concerns propositions that are not relevant to any question S is considering. I used to say in that case belief required credence above 0.5; I now say that S must be willing, at least sometimes, to take *p* for granted.

There are other changes too. I no longer presuppose that questions about what to do just are questions about expected utility. I've stopped focussing exclusively on answers to (conditional) questions, and moved to talking about both answers and ways that questions are answered. And I dropped the requirement that we look at these potentially quite abstruse questions, such as how to answer *Q?* assuming both *A* and *p*. The last two changes offset each other; the reason for including these doubly conditional questions was, in effect, to look at how S was willing to get to answers about questions with more practical import.

There are many reasons, most of them due to perceptive critics of my earlier work, for making these changes. I'll just focus here on the five that have been most significant.

3.7.1 Correctness

Jacob Ross and Mark Schroeder (2014) note that my earlier theory doesn't have a good story about why false beliefs are incorrect.[13] I think that's right. Even if *p* is false, there is nothing necessarily mistaken about either having credence in *p* above 0.5, or in having unconditional preferences match preferences conditional on *p*.

But surely false beliefs are, in a way, incorrect. They may be rational, they may be well-supported, and so on, but still if you believe that *p*, and *p* turns out not to be the case, you got it wrong. There are other mental states that have truth as a correctness condition. Guesses are correct or incorrect, even if there need be nothing at all irrational about making a false guess. Indeed, any mortal who doesn't make false guesses from time to time isn't playing the guessing game well. Not all mental states are like this. Hoping for something that doesn't turn out to happen is unfortunate, but not incorrect. To say that a false belief is incorrect is not to just make the trivial point that it is false. It is also to say that the belief failed to meet one important standard of evaluation for beliefs—correctly representing the world. Credences do not have these correctness conditions, so the relatively simple reduction I proposed of belief to credence must be mistaken.

---

13   Fantl and McGrath (2009) make a similar argument, targeted at Lockean theories of belief more than at my theory. I'll come back to how this is a problem for Lockean theories in Section 8.4.2.

The new theory does not have this problem. Doing dominance reasoning where all of the situations one considers are non-actual is a mistake. It's not a mistake because it will inevitably lead to an irrational decision. Rather, it's a mistake because one draws a conclusion that is not supported by the premises it is based on. Those premises only say that one option is better than another conditional on one or other condition obtaining. That's a bad reason to say the first option is simply better if there is some extra option that might obtain. And whatever does obtain, might obtain.

This way of explaining the incorrectness of false belief suggests a central role for knowledge in norms of beliefs. False beliefs are mistaken because they lead one to treat the actual situation as one that could not obtain, yet the actual situation might obtain. One can make the same mistake by treating a situation that doesn't obtain, but might, as one that could not obtain. Believing something one doesn't know will (typically) lead to doing that.

3.7.2 Impractical Propositions

The second clause in my earlier theory was designed to rule out trivial belief in irrelevant propositions. The first clause on its own has some absurd consequences. Imagine that I'm relaxing by a stream watching the ripples without a care in the world. All of the very few questions that I'm currently interested in have the same answer unconditionally as they do conditional on the Battle of Agincourt having been fought in 1415. So according to clause 1, I believe the Battle of Agincourt was in 1415. That's good, because I do believe that. It's also true that all of the very few questions that I'm currently interested in have the same answer unconditionally as they do conditional on the Battle of Agincourt having been fought in 1416. So if clause 1 was the full theory of belief, then I would also believe that the Battle of Agincourt was in 1416, which I do not.

I added clause 2 to the theory in order to try to fix this problem, but it turned out only to fix a special case. Here's a case it doesn't fix. Let p be the proposition that the next die I roll will land 1, 2, 3 or 4. My credence in that is two-thirds, so it satisfies clause 2. And conditionalising on it doesn't change the answer to any of the very few problems that I'm

interested in while the ripples float down the stream. So I believe p. That's absurd, since I know it is just 2/3 likely. (This objection is also due in important parts to Ross and Schroeder (2014), though my presentation differs from theirs to emphasise just which parts of the objections most worry me.)

The new theory handles this case easily. There is no context where I would simply ignore the possibility that this next die roll will land 5 or 6 for the purposes of doing dominance reasoning. So I don't believe that p, as required.

Is there anything we can rule out on purely probabilistic grounds? It's a little interesting to think this kind of case through. Imagine there is some salient very large number, and it matters what the remainder is when that large number is divided by 1000, or 1000000. Could we get to a point where a choice that is better than some alternative unless that remainder is, say 537, could feel like a dominating choice? I'm not sure whether that would ever happen. It does seem plausible to say that whether such a choice ever feels like a dominating choice correlates with whether we could ever unqualifiedly believe that the remainder is not precisely 537 on purely probabilistic grounds.

3.7.3 Choices with More than Two Options

Consider this variant of the Red-Blue game. As well as the four options Anisa has in the original version of the game, she has a fifth option. This option presents her with a question (as well as the red and blue sentences), and says that if she answers the question correctly, she wins $100. And in this case, the question is, "Who was the first American woman to win an Olympic gold medal?".

Imagine that Anisa just skim-reads the red and blue sentences, and doesn't think about which of them she'd pick, because she knows the answer to this question. It was, she knows, Margaret Abbott. So she promptly gives that answer, and wins $100.

Now she clearly takes an interest in the options Red-True and Blue-True. She has reasons for preferring to answer the question than take one of those two options. And she could give those reasons without any reflection. So Red-True and Blue-True should be in the range of things that we quantify over when thinking about options she is interested in.

Moreover, she has a stable disposition to choose Red-True over Blue-True; I think that stable disposition is a strict preference. That strict preference does not survive conditionalising on the proposition that the Battle of Agincourt was in 1415. So my earlier theory says that even in this revised version of the game, Anisa does not believe that the Battle of Agincourt was in 1415.

This now seems mistaken to me. In any deliberation Anisa does, her regular disposition to take it for granted that the Battle of Agincourt was in 1415 survives. There is a very nearby deliberation where it does not survive, namely the deliberation about whether Red-True or Blue-True is better. But, crucially, she does not have to take an interest in that question in order to take an interest in the two options Red-True and Blue-True. If they are both (clearly) suboptimal options in her current situation, she can simply settle for concluding that they are suboptimal, and leave it at that.

So I think my old theory made it too easy to lose belief in cases where one has to choose between many options. Being interested in some options, because you want to choose the best one of them, does not mean being interested in all questions about preferences between pairs of them. The problem was that I'd been focussing largely on two-way choices, so the distinction between being interested in some choices and being interested in which of those two is better got elided. That distinction matters, and the hybrid pragmatic theory handles it better than my old theory.

3.7.4 Hard Times and Close Calls

In my earlier theory, any practical deliberation was modelled as an inquiry into which option had the highest expected utility. This was wrong for a number of reasons, not least that it gives implausible results in cases involving choices between very similar options. I'll briefly describe one example that illustrates the problem, and the start of how I plan to solve it. It turns out to be rather tricky to get the details right, and I'll come back to this in Section 4.6.1 and again in Chapter 6. The details of the example are new, but it's a very minor modification of a kind of example that is discussed in Matthew McGrath and Brian Kim (2019) and credited to a talk by John Hawthorne "circa 2007". Similar examples

are also discussed by Alex Zweber (2016) and by Charity Anderson and John Hawthorne (2019b), and I'm drawing on their insights in describing this one.

David is doing the weekly groceries. He needs a can of chickpeas, so he walks to where the chickpeas are and looks at the shelf. There are two cans, call them c_1 and c_2, that are equally easy to reach and get from the shelf. Call the actions of taking them t_1 and t_2. David simply assumes, partially on inductive grounds and partially on grounds of what he knows about supermarkets, that neither can has passed its expiry date. While it is wildly implausible that either can has, the probability is not zero. Let e_i be that can i has expired, and assume that $Pr(e_1)$ and $Pr(e_2)$ are low and equal. Call this probability e. Let h be the utility of choosing an unexpired can, and l the utility of choosing an expired can, where obviously $h > l$. Then both t_1 and t_2 have utility $(1-e)h + el$. Conditional on $\neg e_1$, the utility of t_1 is h, which is greater than $(1-e)h + el$ as long as $e > 0$ and $h > l$. So unconditionally, t_1 and t_2 have the same utility, but conditional on $\neg e_1$, they have different utilities. So, according to the theory I used to defend, when David is making this choice, he does not believe, and hence does not know $\neg e_1$. This seems wrong, and there are even worse consequences one can draw my thinking about minor variants of the case.

The key part of my response to this will be distinguishing between the questions *Which can to choose?* and *Which choice of can has maximal expected utility?* If David is thinking about the latter question, then it turns out he really doesn't know $\neg e_1$. That's a somewhat surprising result, and I'll turn to defending it in Chapter 6. But as long as he is focussing solely on the former question, the argument of the previous paragraph doesn't go through.

So the big move here is to move from somewhat quantitative questions, like *Which choice maximises expected utility?*, to practical questions like *What to do?* Once we do that, the problem that Zweber, and Anderson and Hawthorne, raise ceases to be a problem. I don't intend these brief remarks to be a convincing case that I've got a good solution to these problems. Rather, the point is to flag that the theory I'm defending here is distinct from the theory I used to defend, and this gives me some more resources to handle cases like David and the chickpeas.

3.7.5 Updates and Modals

The version of IRT that I defend here gives a big role to conditional attitudes.[14] That's something that it has in common with everything I've written about IRT. I used to have a particular pair of views about how to understand conditional attitudes. In particular, I took the following two claims to be at least close approximations to the truth about conditional attitudes.

1. An attitude conditional on p is (usually) the same as the attitude one would have after updating on p.

2. The way to update on p is to conditionalise.

The first is at best an approximation for familiar reasons. I can think that no one knows whether p is true, and even think that this is true conditional on p. But after updating on p, I will no longer think that. So we have to be a bit careful in applying principle 1; it has counterexamples. Still, it is a useful enough heuristic to work with.

What wasn't originally obvious to me was that there are counterexamples to principle 2 as well. They are more significant for the way IRT should be understood. I used to describe the picture of belief I was defending as the view that to believe something is to have a credence in it that's close enough to 1 for current purposes. That's still a decent heuristic, but it isn't always right. When someone is interested in modal questions, credence 1 might be insufficient for belief. To see how this might be so, it helps to start with some points Thony Gillies (2010) makes about the relationship between modals, conditionals, and updating.

When modal questions are on the table, updating will not be the same as conditionalising. This is shown by the following example. (A similar example is in Kratzer (2012: 94).)

> 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." (Gillies, 2010: 4:13)

---

14   This subsection is based on Weatherson (2016a: §1).

What matters for the purposes of this book is not whether this conditional is true, but whether its truth is consistent with the Ramsey test view of conditionals. And Gillies argues that it is.

> 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. (Gillies, 2010: 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. 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.

All this matters when we are considering modal questions. For example, if we are considering the question *Must q be true?*, then it is plausible that unconditionally the answer is no, and indeed the unconditional probability that q must be true is 0, but that conditional on p, q must be true.

We don't even have to be considering modals directly for this to happen. Assume that actions *A* and *B* have the same outcome conditional on *q*, but *A* is better than *B* in every ¬*q* possibility. Then if we are considering the question *Is A better than B?*, it will matter whether it must be the case that *q*.

Assume that *q* could have probability 1 without it being the case that *q* must be true. (This is controversial, but I'll offer arguments in Sections 8.2 and 8.3 that it is possible.) Then unconditionally, *A* is better than *B*, even though they have the same expected utility. That's because weak dominance is a good principle of practical reasoning: if *A* might be better than *B* and must not be worse, then *A* is better than *B*. But by hypothesis, conditional on *p*, *A* is not better than *B*. So in this case *p* will not be believed; conditional on *p* the question *Is A better than B* gets a different answer to what it gets unconditionally.

Note though that all I said to get this example going is that *p* rules out ¬*q*, and *q* has probability 1. That means *p* could have any probability at all, up to probability 1. So it's possible that conditional on *p*, some relevant questions get different answers to what they get unconditionally, even though *p* has probability 1. So belief can't be a matter of having probability close enough to 1 for practical purposes; sometimes even probability 1 is insufficient.

3.8 Ross and Schroeder's Theory

Jacob Ross and Mark Schroeder (2014) have what looks like, on the surface, a rather different view to mine.[15] They say that to believe *p* is to have a "default reasoning disposition" to use *p* in reasoning. Here's how they describe their 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

---

15   This section is based on Weatherson (2016a: §3).

> 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. (Ross and Schroeder, 2014: 9–10)

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.

The view I'm defending here agrees with them about 1 and 2, though
my theory manifests those characteristics in a quite different way. But
point 3 is a cost of their theory, not a benefit, so it's good that my theory
doesn't accommodate it. (For the record, the theory I put forward in
Weatherson (2005a) did not agree with them on point 2, and I changed
my view because of their arguments.)

   I agree with 1 because, as I've noted a few times above, I think
theoretical interests as well as pragmatic interests matter for the
relationship between credence and belief. I agree with 2 because I
think that whether someone is disposed to use p as a premise matters
to whether they believe p. Let p be some ordinary proposition about
the world that a person believes, such as that the Florida Marlins won
the 2003 World Series. And let q be a lottery proposition that is just as
probable as p. (That is, let q be a lottery proposition such that if the
person were to play the Red-Blue game with p as red and q as blue, they
would be rationally indifferent between the choices.) Then on my theory
the person believes p but not q, and this isn't due to any features of their

credal states. Rather, it is due to their dispositions to use p as a premise in reasoning. (For example, they might use it in figuring out how many World Series were won by National League teams in the 2000s.)

Ross and Schroeder argue, and I basically agree, that interest-relative theories of belief that only focus on practical interests have trouble with folks who use odd techniques in reasoning. This is the lesson of their example of Renzi. The details of that case are unimportant; here's the structure of it. An agent knows that X is better to do if p, and Y is better to do if $\neg p$. They could work out the relative benefit of each option in these two circumstances, and how that interacts with the probability of p to determine which option is best in expectation. They do not in fact do that. Instead, for some proposition q which is not relevant to the case, and very strongly supported by their evidence, they divide into four possibilities: $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$ and $\neg p \wedge \neg q$. They then calculate the expected utility of X and Y given that these are the four possibilities.

This is bad reasoning. Adding this extra division to the possibility space is a waste of time, and increases the chances of making a mistake. They should just use two 'small worlds': p and $\neg p$. The problem we face as theorists is what to say about someone who makes this kind of mistake.

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. Regarding q as open in this way is incompatible with believing it.

I agree. 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.[16]

So far I agree with Ross and Schroeder. The disagreement starts with a principle they endorse, which they call Stability.

---

16 For the record, the theory I defended at the time Ross and Schroeder wrote their paper did not have the resources to make this reply; I've changed my view in light of their arguments.

Stability
A fully rational agent does not change her beliefs purely in virtue of an evidentially irrelevant change in her credences or preferences. (2014: 20)

Stability is motivated by cases like this one.

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 [their name for my old view] 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. (2014: 20)

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. They say 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. 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. Now I agree that rejecting 1 isn't the most intuitive philosophical conclusion one has ever seen. But the alternatives are worse.

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. Still, it is hard to see why this particular instance would fail. Further, 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! Attributing this kind of practical incoherence to Stella is much less plausible than attributing a failure of Stability to her.