

# Knowledge, Bets and Interests

Brian Weatherson

---

## 1 Knowledge in Decision Making

When you pick up a volume like this one, which describes itself as being about ‘knowledge ascriptions’, you probably expect to find it full of papers on epistemology, broadly construed. And you’d probably expect many of those papers to concern themselves with cases where the interests of various parties (ascribers, subjects of the ascriptions, etc.) change radically, and this affects the truth values of various ascriptions. And, at least in this paper, your expectations will be clearly met.

But here’s an interesting contrast. If you’d picked up a volume of papers on ‘belief ascriptions’, you’d expect to find a radically different menu of writers and subjects. You’d expect to find a lot of concern about names and demonstratives, and about how they can be used by people not entirely certain about their denotation. More generally, you’d expect to find less epistemology, and much more mind and language. I haven’t read all the companion papers to mine in this volume, but I bet you won’t find much of that here.

This is perhaps unfortunate, since belief ascriptions and knowledge ascriptions raise at least some similar issues. Consider a kind of contextualism about belief ascriptions, which holds that (L) can be truly uttered in some contexts, but not in others, depending on just what aspects of Lois Lane’s psychology are relevant in the conversation.<sup>1</sup>

(L) Lois Lane believes that Clark Kent is vulnerable to kryptonite.

We could imagine a theorist who says that whether (L) can be uttered truly depends on whether it matters to the conversation that Lois Lane might not recognise Clark Kent when he’s wearing his Superman uniform. And, this theorist might continue, this isn’t because ‘Clark Kent’ is a context-sensitive expression; it is rather because ‘believes’ is context-sensitive. Such a theorist will also, presumably, say that whether (K) can be uttered truly is context-sensitive.

(K) Lois Lane knows that Clark Kent is vulnerable to kryptonite.

And so, our theorist is a kind of contextualist about knowledge ascriptions. But they might agree with approximately none of the motivations for contextualism about knowledge ascriptions put forward by Cohen (1988), DeRose (1995) or Lewis (1996). Rather, they are a contextualist about knowledge ascriptions solely because they are contextualist about belief ascriptions like (L).

Call the position I’ve just described **doxastic contextualism** about knowledge ascriptions. It’s a kind of contextualism all right; it says that whether (K) can be

---

<sup>1</sup>The reflections in the next few paragraphs are inspired by some comments in by Stalnaker (2008), though I don’t want to suggest the theory I’ll discuss is actually Stalnaker’s.

truly uttered is context sensitive, and not because of the context-sensitivity of any term in the ‘that’-clause. But it explains the contextualism solely in terms of the contextualism of belief ascriptions. The more familiar kind of contextualism about knowledge ascriptions we’ll call **non-doxastic contextualism**.

We can make the same kind of division among interest-relative invariantist, or IRI, theories of knowledge ascriptions. Any kind of IRI will say that there are sentences of the form *S knows that p* whose truth depends on the interests, in some sense, of *S*. But we can divide IRI theories up the same way that we divide up contextualist theories.

**Doxastic IRI** Knowledge ascriptions are interest-relative, but their interest-relativity traces solely to the interest-relativity of the corresponding belief ascriptions.

**Non-Doxastic IRI** Knowledge ascriptions are interest-relative, and their interest-relativity goes beyond the interest-relativity of the corresponding belief ascriptions.

In my (2005), I tried to motivate Doxastic IRI. More precisely, I argued for Doxastic IRI about ascriptions of *justified* belief, and hinted that the same arguments would generalise to knowledge ascriptions. I now think those hints were mistaken, and want to defend Non-Doxastic IRI about knowledge ascriptions.<sup>2</sup> My change of heart has been prompted by cases like those Jason Stanley (2005) calls ‘Ignorant High Stakes’ cases.<sup>3</sup> But to see why these cases matter, it will help to start with why I think some kind of IRI must be true. And that story starts with some reflections on the way we teach decision theory.

## 1.1 The Struction of Decision Problems

Professor Dec is teaching introductory decision theory to her undergraduate class. She is trying to introduce the notion of a dominant choice. So she introduces the following problem, with two states,  $S_1$  and  $S_2$ , and two choices,  $C_1$  and  $C_2$  as is normal for introductory problems.

	$S_1$	$S_2$
$C_1$	-\$200	\$1000
$C_2$	-\$100	\$1500

<sup>2</sup>Whether Doxastic or Non-Doxastic IRI is true about justified belief ascriptions turns on some tricky questions about what to say when a subject’s credences are nearly, but not exactly appropriate given her evidence. Space considerations prevent a full discussion of those cases here.

<sup>3</sup>I mean here the case of Coraline, to be discussed in section 3 below. Several people have remarked in conversation that Coraline doesn’t look to them like a case of Ignorant High Stakes. This isn’t surprising; Coraline is better described as being *mistaken* than *ignorant*, and she’s mistaken about odds not stakes. If they’re right, that probably means my argument for Non-Doxastic IRI is less like Stanley’s, and hence more original, than I think it is. So I don’t feel like pressing the point! But I do want to note that I thought the Coraline example was a variation on a theme Stanley originated.

She's hoping that the students will see that  $C_1$  and  $C_2$  are bets, but  $C_2$  is clearly the better bet. If  $S_1$  is actual, then both bets lose, but  $C_2$  loses less money. If  $S_2$  is actual, then both bets win, but  $C_2$  wins more. So  $C_2$  is better. That analysis is clearly wrong if the state is causally dependent on the choice, and controversial if the states are evidentially dependent on the choices. But Professor Dec has not given any reason for the students to think that the states are dependent on the choices in either way, and in fact the students don't worry about that kind of dependence.

That doesn't mean, however, that the students all adopt the analysis that Professor Dec wants them to. One student, Stu, is particularly unwilling to accept that  $C_2$  is better than  $C_1$ . He thinks, on the basis of his experience, that when more than \$1000 is on the line, people aren't as reliable about paying out on bets. So while  $C_1$  is guaranteed to deliver \$1000 if  $S_2$ , if the agent bets on  $C_2$ , she might face some difficulty in collecting on her money.

Given the context, i.e., that they are in an undergraduate decision theory class, it seems that Stu has misunderstood the question that Professor Dec intended to ask. But it is a little harder than it first seems to specify just exactly what Stu's mistake is. It isn't that he thinks Professor Dec has *misdescribed* the situation. It isn't that he thinks it is false that the agent will collect \$1500 if she chooses  $C_2$  and is in  $S_2$ . He just thinks that she *might* not be able to collect it, so the expected payout might really be a little less than \$1500.

Before we try to say just what the misunderstanding between Professor Dec and Stu consists in, let's focus on a simpler problem. Alice is out of town on a holiday, and she faces the following decision choice concerning what to do with a token in her hand.

Choice	Outcome
Put token on table	Win \$1000
Put token in pocket	Win nothing

This looks easy, especially if we've taken Professor Dec's class. Putting the token on the table dominates putting the token in her pocket. It returns \$1000, versus no gain. So she should put the token on the table.

I've left Alice's story fairly schematic; let's fill in some of the details. Alice is on holiday at a casino. It's a fair casino; the probabilities of the outcomes of each of the games is just what you'd expect. And Alice knows this. The table she's standing at is a roulette table. The token is a chip from the casino worth \$1000. Putting the token on the table means placing a bet. As it turns out, it means placing a bet on the roulette wheel landing on 28. If that bet wins she gets her token back and another token of the same value. There are many other bets she could make, but Alice has decided not to make all but one of them. Since her birthday is the 28<sup>th</sup>, she is tempted to put a bet on 28; that's the only bet she is considering. If she makes this bet, the objective chance of her winning is  $1/38$ , and she knows this. As a matter of fact she will win, but she doesn't know this. (This is why the description in the table I presented above is truthful, though frightfully misleading.) As you can see, the odds on this bet are

terrible. She should have a chance of winning around  $1/2$  to justify placing this bet.<sup>4</sup> So the above table, which makes it look like placing the bet is the dominant, and hence rational, option, is misleading.

Just how is the table misleading though? It isn't because what it says is false. If Alice puts the token on the table she wins \$1000; and if she doesn't, she stays where she is. It isn't, or isn't just, that Alice doesn't believe the table reflects what will happen if she places the bet. As it turns out, Alice is smart, so she doesn't form beliefs about chance events like roulette wheels. But even if she did, that wouldn't change how misleading the table is. The table suggests that it is rational for Alice to put the token on the table. In fact, that is irrational. And it would still be irrational if Alice believes, *irrationally*, that the wheel will land on 28.

A better suggestion is that the table is misleading because Alice doesn't *know* that it accurately depicts the choice she faced. If she did know that these were the outcomes to putting the token on the table versus in her pocket, it seems it would be rationally compelling for her to put it on the table. If we take it as tacit in a presentation of a decision problem that the agent knows that the table accurately depicts the outcomes of various choices in different states, then we can tell a plausible story about what the miscommunication between Professor Dec and Stu was. Stu was assuming that if the agent wins \$1500, she might not be able to easily collect. That is, he was assuming that the agent does not know that she'll get \$1500 if she chooses  $C_2$  and is in state  $S_2$ . Professor Dec, if she's anything like other decision theory professors, will have assumed that the agent did know exactly that.

As we've seen, the standard presentation of a decision problem presupposes not just that the table states what will happen, but the agent stands in some special doxastic relationship to that information. Could that relationship be weaker than knowledge? It's true that it is hard to come up with clear counterexamples to the suggestion that the relationship is merely justified true belief. But I think it is somewhat implausible to hold that the standard presentation of an example merely presupposes that the agent has a justified true belief that the table is correct, and does not in addition know that the table is correct.

My reasons for thinking this are similar to one of the reasons Timothy Williamson (Williamson, 2000, Ch. 9) gives for doubting that one's evidence is all that one justifiably truly believes. To put the point in Lewisian terms, it seems that knowledge is a much more *natural* relation than justified true belief. And when ascribing contents, especially contents of tacitly held beliefs, we should strongly prefer to ascribe more rather than less natural contents.

I'm here retracting some things I said a few years ago in a paper on philosophical methodology (Weatherson, 2003). There I argued that identifying knowledge with justified true belief would give us a theory on which knowledge was more natural than a theory on which we didn't identify knowledge with any other epistemic property. I now think that is wrong for a couple of reasons. First, although it's true (as I say in the earlier paper) that knowledge can't be primitive or perfectly natural, this

<sup>4</sup>Assuming Alice's utility curve for money curves downwards, she should be looking for a slightly higher chance of winning than  $1/2$  to place the bet, but that level of detail isn't relevant to the story we're telling here.

doesn't make it less natural than justification, which is also far from a fundamental feature of reality. Indeed, given how usual it is for languages to have a simple representation of knowledge, we have some evidence that it is very natural for a term from a special science. Second, I think in the earlier paper I didn't fully appreciate the point (there attributed to Peter Klein) that the Gettier cases show that the property of being a justified true belief is not particularly natural. In general, when  $F$  and  $G$  are somewhat natural properties, then so is the property of being  $F \wedge G$ . But there are exceptions, especially in cases where these are properties that a whole can have in virtue of a part having the property. In those cases, a whole that has an  $F$  part and a  $G$  part will be  $F \wedge G$ , but this won't reflect any distinctive property of the whole. And one of the things the Gettier cases show is that the properties of *being justified* and *being true*, as applied to belief, fit this pattern.<sup>5</sup>

So the 'special doxastic relationship' is not weaker than knowledge. Could it be stronger? Could it be, for example, that the relationship is certainty, or some kind of iterated knowledge? Plausibly in some game-theoretic settings it is stronger – it involves not just knowing that the table is accurate, but knowing that the other player knows the table is accurate. In some cases, the standard treatment of games will require positing even more iterations of knowledge. For convenience, it is sometimes explicitly stated that iterations continue indefinitely, so each party knows the table is correct, and knows each party knows this, and knows each party knows that, and knows each party knows *that*, and so on. An early example of this in philosophy is in the work by David Lewis (1969) on convention. But it is usually acknowledged (again in a tradition extending back at least to Lewis) that only the first few iterations are actually needed in any problem, and it seems a mistake to attribute more iterations than are actually used in deriving solutions to any particular game.

The reason that would be a mistake is that we want game theory, and decision theory, to be applicable to real-life situations. There is very little that we know, and know that we know, and know we know we know, and so on indefinitely (Williamson, 2000, Ch. 4). There is, perhaps, even less that we are certain of. If we only could say that a person is playing a particular game when they stand in these very strong relationships to the parameters of the game, then people will almost never be playing any games of interest. Since game theory, and decision theory, are not meant to be that impractical, I conclude that the 'special doxastic relationship' cannot be that strong. It could be that in some games, the special relationship will involve a few iterations of knowledge, but in decision problems, where the epistemic states of others are irrelevant, even that is unnecessary, and simple knowledge seems sufficient.

It might be argued here that we shouldn't expect to apply decision theory directly to real-life problems, but only to idealised versions of them, so it would be acceptable

<sup>5</sup>Note that even if you think that philosophers are generally too quick to move from instinctive reactions to the Gettier case to abandoning the justified true belief theory of knowledge, this point holds up. What is important here is that on sufficient reflection, the Gettier cases show that some justified true beliefs are not knowledge, and that the cases in question also show that being a justified true belief is not a particularly natural or unified property. So the point I've been making in the last few paragraphs is independent of the point I wanted to stress in "What Good are Counterexamples?", namely, that philosophers in some areas (especially epistemology) are insufficiently reformist in their attitude towards our intuitive reactions to cases.

to, for instance, require that the things we put in the table are, say, things that have probability exactly 1. In real life, virtually nothing has probability 1. In an idealisation, many things do. But to argue this way seems to involve using ‘idealisation’ in an unnatural sense. There is a sense in which, whenever we treat something with non-maximal probability as simply given in a decision problem that we’re ignoring, or abstracting away from, some complication. But we aren’t *idealising*. On the contrary, we’re modelling the agent as if they were irrationally certain in some things which are merely very very probable.

So it’s better to say that any application of decision theory to a real-life problem will involve ignoring certain (counterfactual) logical or metaphysical possibilities in which the decision table is not actually true. But not any old abstraction will do. We can’t ignore just anything, at least not if we want a good model. Which abstractions are acceptable? I have an answer to this: we can abstract away from any possibility in which something the agent actually knows is false. I don’t have a knock-down argument that this is the best of all possible abstractions, but nor do I know of any alternative answer to the question which abstractions are acceptable which is nearly as plausible.

In part that is because it is plausible that the ‘special doxastic relationship’ should be a fairly simple, natural relationship. And it seems that any simple, natural relationship weaker than knowledge will be so weak that when we plug it into our decision theory, it will say that Alice should do clearly irrational things in one or other of the cases we described above. And it seems that any simple, natural relationship stronger than knowledge will be so strong that it makes decision theory or game theory impractical.

I also cheated a little in making this argument. When I described Alice in the casino, I made a few explicit comments about her information states. And every time, I said that she *knew* various propositions. It seemed plausible at the time that this is enough to think those propositions should be added to the table. That’s some evidence against the idea that more than knowledge, perhaps iterated knowledge or certainty, is needed before we add propositions to the decision table.

## 1.2 From Decision Theory to Interest-Relativity

This way of thinking about decision problems offers a new perspective on the issue of whether we should always be prepared to bet on what we know.<sup>6</sup> To focus intuitions, let’s take a concrete case. Barry is sitting in his apartment one evening when he hears a musician performing in the park outside. The musician, call her Beth, is one of Barry’s favourite musicians, so the music is familiar to Barry. Barry is excited that Beth is performing in his neighbourhood, and he decides to hurry out to see the show. As he prepares to leave, a genie appears and offers him a bet. If he takes the bet, and the musician is Beth, then the genie gives Barry ten dollars. On the other hand, if the musician is not Beth, he will be tortured in the fires of hell for a millenium. Let’s put Barry’s options in table form.

<sup>6</sup>This issue is of course central to the plotline in Hawthorne (2004).

	<b>Musician is Beth</b>	<b>Musician is not Beth</b>
<b>Take Bet</b>	Win \$10	1000 years of torture
<b>Decline Bet</b>	Status quo	Status quo

Intuitively, it is extremely irrational for Barry to take the bet. People do make mistakes about identifying musicians, even very familiar musicians, by the strains of music that drift up from a park. It's not worth risking a millenium of torture for \$10.

But it also seems that we've misstated the table. Before the genie showed up, it seemed clear that Barry knew that the musician was Beth. That was why he went out to see her perform. (If you don't think this is true, make the sounds from the park clearer, or make it that Barry had some prior evidence that Beth was performing which the sounds from the park remind him of. It shouldn't be too hard to come up with an evidential base such that (a) in normal circumstances we'd say Barry knew who was performing, but (b) he shouldn't take this genie's bet.) Now our decision tables should reflect the knowledge of the agent making the decision. If Barry knows that the musician is Beth, then the second column is one he knows will not obtain. Including it is like including a column for what will happen if the genie is lying about the consequences of taking or declining the bet. So let's write the table in the standard form.

	<b>Musician is Beth</b>
<b>Take Bet</b>	Win \$10
<b>Decline Bet</b>	Status quo

And it is clear what Barry's decision should be in this situation. Taking the bet dominates declining it, and Barry should take dominating options.

What has happened? It is incredibly clear that Barry should decline the bet, yet here we have an argument that he should take the bet. If you accept that the bet should be declined, then there are three options available it seems to me.

1. Barry never knew that the musician was Beth.
2. Barry did know that the musician was Beth, but this knowledge was destroyed by the genie's offer of the bet.
3. States of the world that are known not to obtain should still be represented in decision problems, so taking the bet is not a dominating option.

The first option is basically a form of scepticism. If the take-away message from the above discussion is that Barry doesn't know the musician is Beth, we can mount a similar argument to show that he knows next to nothing.<sup>7</sup> And the third option would send us back into the problems about interpreting and applying decision theory that we spend the first few pages trying to get out of.

<sup>7</sup>The idea that interest-relativity is a way of fending off scepticism is a very prominent theme in Fantl and McGrath (2009).

So it seems that the best solution here, or perhaps the least bad solution, is to accept that knowledge is interest-relative. Barry did know that the musician was Beth, but the genie's offer destroyed that knowledge.

The argument here bears more than a passing resemblance to the arguments in favour of interest-relativity that are made by Hawthorne, Stanley and Fantl and McGrath. But I think the focus on decision theory shows how we get to interest-relativity with slightly weaker premises than they are using.<sup>8</sup> In particular, the only premises I've used to derive an interest-relative conclusion are:

1. Before the genie showed up, Barry knew the musician was Beth.
2. It's rationally permissible, *in cases like Barry's*, to take dominating options.
3. It's always right to model decision problems by including what the agent knows in the 'framework', i.e., the specification of the options and payouts.
4. It is rationally impermissible for Barry to take the genie's offered bet.

The second premise there is *much* weaker than the principles linking knowledge and action defended in previous arguments for interest-relativity. It isn't the claim that one can always act on what one knows, or that one can only act on what one knows, or that knowledge always (or only) provides reason to act. It's just the claim that in one very specific type of situation, in particular when one has to make a relatively simple bet, which affects nobody but the person making the bet, it's rationally permissible to take a dominating option. In conjunction with the third premise, it entails that *in those kind of cases*, the fact that one knows taking the bet will lead to a better outcome suffices for making acceptance of the bet rationally permissible. It doesn't say anything about what else might or might not make acceptance rationally permissible. It doesn't say anything about what suffices for rational permissibility in other kinds of cases, such as cases where someone else's interests are at stake, or where taking the bet might violate a deontological constraint, or any other way in which real-life choices differ from the simplest decision problems.<sup>9</sup> It doesn't say anything about any other kind of permissibility, e.g., moral permissibility. But it doesn't need to, because we're only in the business of proving that there is *some* interest-relativity to knowledge, and an assumption about practical rationality in some range of cases suffices to prove that.

The case of Barry and Beth also bears some relationship to one of the kinds of case that have motivated contextualism about knowledge. Indeed, it has been widely noted in the literature on interest-relativity that interest-relativity can explain away many of the puzzles that motivate contextualism. And there are difficulties that face any contextualist theory (Weatherson, 2006). So I prefer an *invariantist* form

<sup>8</sup>As they make clear in Hawthorne and Stanley (2008), Hawthorne and Stanley are interested in defending relatively strong premises linking knowledge and action independently of the argument for the interest-relativity of knowledge. What I'm doing here is showing how that conclusion does not rest on anything nearly as strong as the principles they believe, and so there is plenty of space to disagree with their general principles, but accept interest-relativity.

<sup>9</sup>I have more to say about those cases in section 2.2.



of interest-relativity about knowledge. That is, my view is a form of interest-relative-invariantism, or IRI.<sup>10</sup>

Now everything I've said here leaves it open whether the interest-relativity of knowledge is a natural and intuitive theory, or whether it is a somewhat unhappy concession to difficulties that the case of Barry and Beth raise. I think the former is correct, and interest-relativity is fairly plausible on its own merits, but it would be consistent with my broader conclusions to say that in fact the interest-relative theory of knowledge is very implausible and counterintuitive. If we said that, we could still justify the interest-relative theory by noting that we have on our hands here a paradoxical situation, and any option will be somewhat implausible. This consideration has a bearing on how we should think about the role of intuitions about cases, or principles, in arguments that knowledge is interest-relative. Several critics of the view have argued that the view is counter-intuitive, or that it doesn't accord with the reactions of non-expert judges.<sup>11</sup> In a companion paper, "Defending Interest-Relative Invariantism", I note that those arguments usually misconstrue what the consequences of interest-relative theories of knowledge are. But even if they don't, I don't think there's any quick argument that if interest-relativity is counter-intuitive, it is false. After all, the only alternatives that seem to be open here are very counter-intuitive.

Finally, it's worth noting that if Barry is rational, he'll stop (fully) believing that the musician is Beth once the genie makes the offer. Assuming the genie allows this, it would be very natural for Barry to try to acquire more information about the singer. He might walk over to the window to see if he can see who is performing in the park. So this case leaves it open whether the interest-relativity of knowledge can be explained fully by the interest-relativity of belief. I used to think it could be; I no longer think that. To see why this is so, it's worth rehearsing how the interest-relative theory of belief runs.

## 2 The Interest-Relativity of Belief

### 2.1 Interests and Functional Roles

The previous section was largely devoted to proving an existential claim: there is *some* interest-relativity to knowledge. Or, if you prefer, it proved a negative claim: the best theory of knowledge is *not* interest-neutral. But this negative conclusion invites a philosophical challenge: what is the best explanation of the interest-relativity of knowledge? My answer is in two parts. Part of the interest-relativity of knowledge comes from the interest-relativity of belief, and part of it comes from the fact that interests generate certain kinds of doxastic defeaters.

We can see that belief is interest-relative by seeing that the best functionalist theory of belief is interest-relative. Since the best functionalist theory of belief is the true theory of belief, that means belief is interest-relative. That last sentence assumed a non-trivial premise, namely that functionalism about belief is true. I'm not going to

<sup>10</sup>This is obviously not a full argument against contextualism; that would require a much longer paper than this.

<sup>11</sup>See, for instance, Blome-Tillmann (2009), or Feltz and Zarpentine (forthcoming).

argue for that, since the argument would take at least a book. (The book in question might look a lot like Braddon-Mitchell and Jackson (2007).) But I am going to, in this section, argue for what I said in the first sentence of this paragraph, namely that the best functionalist account of belief is interest-relative.

In my (2005), I suggested that was all of the explanation of the interest-relativity of knowledge. I was wrong, and section 3 of this paper will show why I was wrong. Interests also generate certain kinds of doxastic defeaters, and they enter independently into the explanation of why knowledge is interest-relative. Very roughly, the idea is that *S* doesn't know *p* if *S*'s belief that *p* does not sufficiently cohere with the rest of what she should believe and does believe. If this coherence constraint fails, there is a doxastic defeater. When is some incoherence too much incoherence for knowledge? That turns out to be interest-relative, in cases like the case of Coraline described in the next section. But we're getting ahead of ourselves, the first task is to link functionalism about belief and the interest-relativity of belief.

We start not with the functional characterisation of belief, but with something closeby, the functional characterisation of credence. Frank Ramsey (1926) provides a clear statement of one of the key functional roles of credences; their connection to action. Of course, Ramsey did not take himself to be providing one component of the functional theory of credence. He took himself to be providing a behaviourist/operationalist reduction of credences to dispositions. But we do not have to share Ramsey's metaphysics to use his key ideas. Those ideas include that it's distinctively *betting* dispositions that are crucial to the account of credence, and that all sorts of actions in everyday life constitute bets.

The connection to betting behaviour lives on today most prominently in the work on 'representation theorems'.<sup>12</sup> What a representation theorem shows is that for any agent whose pairwise preferences satisfy some structural constraints, there is a probability function and a utility function such that the agent prefers bet *X* to bet *Y* just in case the expected utility of *X* (given that probability and utility function) is greater than that of *Y*. Moreover, the probability function is unique (and the utility function is unique up to positive affine transformations). Given that, it might seem plausible to identify the agent's credence with this probability function, and the agent's (relative) values with this utility function.

Contemporary functionalism goes along with much, but not quite all, of this picture. The betting preferences are an important part of the functional role of a credence; indeed, they just are the output conditions. But there are two other parts to a functional role: an input condition and a set of internal connections. So the functionalist thinks that the betting dispositions are not quite sufficient for having credences. A pre-programmed automaton might have dispositions to accept (or at least move as if accepting) various bets, but this will not be enough for credences (Braddon-Mitchell and Jackson, 2007). So when we're considering whether someone is in a particular credal state, we have to consider not just their actions, and their disposition to act, but the connections between the alleged credal state and other states.

<sup>12</sup>See Maher (1993) for the most developed account in recent times.

The same will be true of belief. Part of what it is to believe  $p$  is to act as if  $p$  is true. Another part is to have other mental states that make sense in light of  $p$ . This isn't meant to rule out the possibility of inconsistent beliefs. As David Lewis (1982) points out, it is easy to have inconsistent beliefs if we don't integrate our beliefs fully. What it is meant to rule out is the possibility of a belief that is not at all integrated into the agent's cognitive system. If the agent believes  $q$ , but refuses to infer  $p \wedge q$ , and believes  $p \rightarrow r$ , but refuses to infer  $r$ , and so on for enough other beliefs, she doesn't really believe  $p$ .

When writers start to think about the connection between belief and credence, they run into the following problem fairly quickly. The alleged possibility where  $S$  believes  $p$ , and doesn't believe  $q$ , but her credence in  $q$  is higher than her credence in  $p$  strikes many theorists as excessively odd.<sup>13</sup> That suggests the so-called 'threshold view' of belief, that belief is simply credence above a threshold. It is also odd to say that a rational, reflective agent could believe  $p$ , believe  $q$ , yet take it as an open question whether  $p \wedge q$  is true, refusing to believe or disbelieve it. Finally, it seems we can believe things to which we don't give credence 1. In the case of Barry and Beth from section 1, for example, before the genie comes in, it seems Barry does believe the musician is Beth. But he doesn't have credence 1 in this, since having credence 1 means being disposed to make a bet at any odds. And we can't have all three of these ideas. That is, we can't accept the 'threshold view', the closure of belief under conjunction, and the possibility that propositions with credence less than one are believed.

We can raise the same kind of problem by looking directly at functional roles. A key functional role of credences is that if an agent has credence  $x$  in  $p$  she should be prepared to buy a bet that returns 1 util if  $p$ , and 0 utils otherwise, iff the price is no greater than  $x$  utils. A key functional role of belief is that if an agent believes  $p$ , and recognises that  $\phi$  is the best thing to do given  $p$ , then she'll do  $\phi$ . Given  $p$ , it's worth paying any price up to 1 util for a bet that pays 1 util if  $p$ . So believing  $p$  seems to mean being in a functional state that is like having credence 1 in  $p$ . But as we argued in the previous paragraph, it is wrong to identify belief with credence 1.

If we spell out more carefully what the functional states of credence and belief are, a loophole emerges in the argument that belief implies credence 1. The interest-relative theory of belief exploits that loophole. What's the difference, in functional terms, between having credence  $x$  in  $p$ , and having credence  $x + \varepsilon$  in  $p$ ? Well, think again about the bet that pays 1 util if  $p$ , and 0 utils otherwise. And imagine that bet is offered for  $x + \varepsilon/2$  utils. The person whose credence is  $x$  will decline the offer; the person whose credence is  $x + \varepsilon$  will accept it. Now it will usually be that no such bet is on offer.<sup>14</sup> No matter; as long as one agent is *disposed* to accept the offer, and the

<sup>13</sup>I'm going to argue below that cases like that of Barry and Beth suggest that in practice this isn't nearly as odd as it first seems.

<sup>14</sup>There are exceptions, especially in cases where  $p$  concerns something significant to financial markets, and the agent trades financial products. If you work through the theory that I'm about to lay out, one consequence is that such agents should have very few unconditional beliefs about financially-sensitive information, just higher and lower credences. I think that's actually quite a nice outcome, but I'm not going to rely on that in the argument for the view.

other agent is not, that suffices for a difference in credence.

The upshot of that is that differences in credences might be, indeed usually will be, constituted by differences in dispositions concerning how to act in choice situations far removed from actuality. I'm not usually in a position of having to accept or decline a chance to buy a bet for 0.9932 utils that the local coffee shop is currently open. Yet whether I would accept or decline such a bet matters to whether my credence that the coffee shop is open is 0.9931 or 0.9933. This isn't a problem with the standard picture of how credences work. It's just an observation that the high level of detail embedded in the picture relies on taking the constituents of mental states to involve many dispositions.

It isn't clear that belief should be defined in terms of the same kind of dispositions involving better behaviour in remote possibilities. It's true that if I believe that  $p$ , and I'm rational enough, I'll act as if  $p$  is true. Is it also true that if I believe  $p$ , I'm disposed to act as if  $p$  is true no matter what choices are placed in front of me? I don't see any reason to say yes, and there are a few reasons to say no. As we say in the case of Barry and Beth, Barry can believe that  $p$ , but be disposed to *lose that belief* rather than act on it if odd choices, like that presented by the genie, emerge.

This suggests the key difference between belief and credence 1. For a rational agent, a credence of 1 in  $p$  means that the agent is disposed to answer a wide range of questions the same way she would answer that question conditional on  $p$ . That follows from the fact that these four principles are trivial theorems of the orthodox theory of expected utility.<sup>15</sup>

**C1AP** For all  $q, x$ , if  $\Pr(p) = 1$  then  $\Pr(q) = x$  iff  $\Pr(q|p) = x$ .

**C1CP** For all  $q, r$ , if  $\Pr(p) = 1$  then  $\Pr(q) \geq \Pr(r)$  iff  $\Pr(q|p) \geq \Pr(r|p)$ .

**C1AU** For all  $\phi, x$ , if  $\Pr(p) = 1$  then  $U(\phi) = x$  iff  $U(\phi|p) = x$ .

**C1CP** For all  $\phi, \psi$ , if  $\Pr(p) = 1$  then  $U(\phi) \geq U(\psi)$  iff  $U(\phi|p) \geq U(\psi|p)$ .

In the last two lines, I use  $U(\phi)$  to denote the expected utility of  $\phi$ , and  $U(\phi|p)$  to denote the expected utility of  $\phi$  conditional on  $p$ . It's often easier to write this as simply  $U(\phi \wedge p)$ , since the utility of  $\phi$  conditional on  $p$  just is the utility of doing  $\phi$  in a world where  $p$  is true. That is, it is the utility of  $\phi \wedge p$  being realised. But we get a nicer symmetry between the probabilistic principles and the utility principles if we use the explicitly conditional notation for each.

If we make the standard kinds of assumptions in orthodox decision theory, i.e., assume at least some form of probabilism and consequentialism<sup>16</sup>, then the agent will answer each of these questions the same way simpliciter and conditional on  $p$ .

- How probable is  $q$ ?
- Is  $q$  or  $r$  more probable?
- How good an idea is it to do  $\phi$ ?

<sup>15</sup>The presentation in this section, as in Weatherson (2005), assumes at least a weak form of consequentialism. This was arguably a weakness of the earlier paper. We'll return to the issue of what happens in cases where the agent doesn't, and perhaps shouldn't, maximise expected utility, at the end of the section.

<sup>16</sup>I mean here consequentialism in roughly the sense used by Hammond (1988).

- Is it better to do  $\phi$  or  $\psi$ ?

Each of those questions is schematic. As in the more technical versions given above, they quantify over propositions and actions, albeit tacitly in the case of these versions. And these quantifiers have a very large domain. The standard theory is that an agent whose credence in  $p$  is 1 will have the same credence in  $q$  as in  $q$  given  $p$  for any  $q$  whatsoever.

Now in one sense, exactly the same things are true if the agent believes  $p$ . If one is wondering whether  $q$  or  $r$  is more probable, and one believes  $p$ , then the fact that  $p$  will be taken as a given in structuring this inquiry. So conditionalising on  $p$  should not change the answer to the question. And the same goes for any actions  $\phi$  and  $\psi$  that the agent is choosing between. But this isn't true unrestrictedly. That's what we saw in the case of Barry and Beth. Which choices are on the table will change which things the agent will take as given, will use to structure inquiry, will believe. If we return to the technical version of our four questions, we might put all this the following way.

**BAP** For all relevant  $q, x$ , if  $p$  is believed then  $\Pr(q) = x$  iff  $\Pr(q|p) = x$ .

**BCP** For all relevant  $q, r$ , if  $p$  is believed then  $\Pr(q) \geq \Pr(r)$  iff  $\Pr(q|p) \geq \Pr(r|p)$ .

**BAU** For all relevant  $\phi, x$ , if  $p$  is believed then  $U(\phi) = x$  iff  $U(\phi|p) = x$ .

**BCP** For all relevant  $\phi, \psi$ , if  $p$  is believed then  $U(\phi) \geq U(\psi)$  iff  $U(\phi|p) \geq U(\psi|p)$ .

This is where interests, theoretical or practical, matter. The functional definition of belief looks a lot like the functional definition of credence 1. But there's one key difference. Both definitions involve quantifiers: quantifying over propositions, actions and values. In the definition of credence 1, those quantifiers are (largely) unrestricted. In the definition of belief, those quantifiers are tightly restricted, and the restrictions are in terms of the interests, practical and theoretical, of the agent.

In the earlier paper I went into a lot of detail about just what 'relevant' means in this context, and I won't repeat that here. But I will say a little about one point I didn't sufficiently stress in that paper: the importance of the restriction of **BAP** and **BAU** to *relevant* values of  $x$ . This lets us have the following nice consequence.

Charlie is trying to figure out exactly what the probability of  $p$  is. That is, for any  $x \in [0, 1]$ , whether  $\Pr(p) = x$  is a relevant question. Now Charlie is well aware that  $\Pr(p|p) = 1$ . So unless  $\Pr(p) = 1$ , Charlie will give a different answer to the questions *How probable is  $p$ ?* and *Given  $p$ , how probable is  $p$ ?* So unless Charlie holds that  $\Pr(p)$  is 1, she won't count as believing that  $p$ . One consequence of this is that Charlie can't reason, "The probability of  $p$  is exactly 0.978, so  $p$ ." That's all to the good, since that looks like bad reasoning. And it looks like bad reasoning even though in some circumstances Charlie can rationally believe propositions that she (rationally) gives credence 0.978 to.

But note that the reasoning in the previous paragraph assumes that every question of the form *Is the probability of  $p$  equal to  $x$ ?* is relevant. In practice, fewer questions than that will be relevant. Let's say that the only questions relevant to Charlie are of the form *What is the probability of  $p$  to one decimal place?*. And assume that no

other questions become relevant in the course of her inquiry into this question.<sup>17</sup> Charlie decides that to the first decimal place,  $\Pr(p) = 1.0$ , i.e.,  $\Pr(p) > 0.95$ . That is compatible with simply believing that  $p$ . And that seems right; if for practical purposes, the probability of  $p$  is indistinguishable from 1, then the agent is confident enough in  $p$  to believe it.

## 2.2 Two Caveats

The theory sketched so far seems to me right in the vast majority of cases. It fits in well with a broadly functionalist view of the mind, and it handles difficult cases, like that of Kate, nicely. But it needs to be supplemented and clarified a little to handle some other difficult cases. In this section I'm going to supplement the theory a little to handle what I call 'impractical propositions', and say a little about morally loaded action.

Jones has a false geographic belief: he believes that Los Angeles is west of Reno, Nevada.<sup>18</sup> This isn't because he's ever thought about the question. Rather, he's just disposed to say "Of course" if someone asks, "Is Los Angeles west of Reno?" That disposition has never been triggered, because no one's ever bothered to ask him this. Call the proposition that Los Angeles is west of Reno  $p$ .

The theory given so far will get the right result here: Jones does believe that  $p$ . But it gets the right answer for an odd reason. Jones, it turns out, has very little interest in American geography right now. He's a schoolboy in St Andrews, Scotland, getting ready for school and worried about missing his schoolbus. There's no inquiry he's currently engaged in for which  $p$  is even close to relevant. So conditionalising on  $p$  doesn't change the answer to any inquiry he's engaged in, but that would be true no matter what his credence in  $p$  is.

There's an immediate problem here. Jones believes  $p$ , since conditionalising on  $p$  doesn't change the answer to any relevant inquiry. But for the very same reason, conditionalising on  $\neg p$  doesn't change the answer to any relevant inquiry. It seems our theory has the bizarre result that Jones believes  $\neg p$  as well. That is both wrong and unfair. We end up attributing inconsistent beliefs to Jones simply because he's a harried schoolboy who isn't currently concerned with the finer points of the geography of the American southwest.

Here's a way out of this problem in four relatively easy steps. First, we say that which questions are relevant questions is not just relative to the agent's interests, but also relevant to the proposition being considered. A question may be relevant relative to  $p$ , but not relative to  $q$ . Second, we say that relative to  $p$ , the question of whether to believe  $p$  is a relevant question. Third, we say that an agent only prefers believing  $p$  to not believing it if their credence in  $p$  is greater than their credence in  $\neg p$ , i.e., if their credence in  $p$  is greater than  $1/2$ . Finally, we say that when the issue is whether the subject believes that  $p$ , the question of whether to believe  $p$  is not just a relevant

<sup>17</sup>This is probably somewhat unrealistic. It's hard to think about whether  $\Pr(p)$  is closer to 0.7 or 0.8 without raising to salience questions about, for example, what the second decimal place in  $\Pr(p)$  is. This is worth bearing in mind when coming up with intuitions about the cases in this paragraph.

<sup>18</sup>I'm borrowing this example from Fred Dretske, who uses it to make some interesting points about dispositional belief.

question on its own, but it stays being a relevant question conditional on any  $q$  that is relevant to the subject. In the earlier paper (Weatherson, 2005) I argue that this solves the problem raised by impractical propositions in a smooth and principled way.

That's the first caveat. The second is one that isn't discussed in the earlier paper. If the agent is merely trying to get the best outcome for themselves, then it makes sense to represent them as a utility maximiser. And within orthodox decision theory, it is easy enough to talk about, and reason about, conditional utilities. That's important, because conditional utilities play an important role in the theory of belief offered here. But if the agent faces moral constraints on her decision, it isn't always so easy to think about conditional utilities.

When agents have to make decisions that might involve them causing harm to others if certain propositions turn out to be true, then I think it is best to supplement orthodox decision theory with an extra assumption. The assumption is, roughly, that for choices that may harm others, expected value is absolute value. It's easiest to see what this means using a simple case of three-way choice. The kind of example I'm considering here has been used for (slightly) different purposes by Frank Jackson (1991).

The agent has to do  $\varphi$  or  $\psi$ . Failure to do either of these will lead to disaster, and is clearly unacceptable. Either  $\varphi$  or  $\psi$  will avert the disaster, but one of them will be moderately harmful and the other one will not. The agent has time before the disaster to find out which of  $\varphi$  and  $\psi$  is harmful and which is not for a nominal cost. Right now, her credence that  $\varphi$  is the harmful one is, quite reasonably,  $1/2$ . So the agent has three choices:

- Do  $\varphi$ ;
- Do  $\psi$ ; or
- Wait and find out which one is not harmful, and do it.

We'll assume that other choices, like letting the disaster happen, or finding out which one is harmful and doing it, are simply out of consideration. In any case, they are clearly dominated options, so the agent shouldn't do them. Let  $p$  be the proposition that  $\varphi$  is the harmful one. Then if we assume the harm in question has a disutility of 10, and the disutility of waiting to act until we know which is the harmful one is 1, the values of the possible outcomes are as follows:

	$p$	$\neg p$
Do $\varphi$	-10	0
Do $\psi$	0	-10
Find out which is harmful	-1	-1

Given that  $Pr(p) = 1/2$ , it's easy to compute that the expected value of doing either  $\varphi$  or  $\psi$  is -5, while the expected value of finding out which is harmful is -1, so the agent should find out which thing is to be done before acting. So far most consequentialists

would agree, and so probably would most non-consequentialists for most ways of fleshing out the abstract example I've described.<sup>19</sup>

But most consequentialists would also say something else about the example that I think is not exactly true. Just focus on the column in the table above where  $p$  is true. In that column, the highest value, 0, is alongside the action  $Do\ \psi$ . So you might think that conditional on  $p$ , the agent should do  $\psi$ . That is, you might think the conditional expected value of doing  $\psi$ , conditional on  $p$  being true, is 0, and that's higher than the conditional expected value of any other act, conditional on  $p$ . If you thought that, you'd certainly be in agreement with the orthodox decision-theoretic treatment of this problem.

In the abstract statement of the situation above, I said that one of the options would be *harmful*, but I didn't say who it would be harmful to. I think this matters. I think what I called the orthodox treatment of the situation is correct when the harm accrues to the person making the decision. But when the harm accrues to another person, particularly when it accrues to a person that the agent has a duty of care towards, then I think the orthodox treatment isn't quite right.

My reasons for this go back to Jackson's original discussion of the puzzle. Let the agent be a doctor, the actions  $\varphi$  and  $\psi$  be her prescribing different medication to a patient, and the harm a severe allergic reaction that the patient will have to one of the medications. Assume that she can run a test that will tell her which medication the patient is allergic to, but the test will take a day. Assume that the patient will die in a month without either medication; that's the disaster that must be averted. And assume that the patient is in some discomfort that either medication would relieve; that's the small cost of finding out which medication is risk. Assume finally that there is no chance the patient will die in the day it takes to run the test, so the cost of running the test is really nominal.

A good doctor in that situation will find out which medication the patient is allergic to before ascribing either medicine. It would be *reckless* to ascribe a medicine that is unnecessary and that the patient might be allergic to. It is worse than reckless if the patient is actually allergic to the medicine prescribed, and the doctor harms the patient. But even if she's lucky and prescribes the 'right' medication, the recklessness remains. It was still, it seems, the wrong thing for her to do.

All of that is in Jackson's discussion of the case, though I'm not sure he'd agree with the way I'm about to incorporate these ideas into the formal decision theory. Even under the assumption that  $p$ , prescribing  $\psi$  is still wrong, because it is reckless. That should be incorporated into the values we ascribe to different actions in different circumstances. The way I do it is to associate the value of each action, in each circumstance, with its actual expected value. So the decision table for the doctor's decision looks something like this.

<sup>19</sup>Some consequentialists say that what the agent should do depends on whether  $p$  is true. If  $p$  is true, she should do  $\psi$ , and if  $p$  is false she should do  $\varphi$ . As we'll see, I have reasons for thinking this is rather radically wrong.



	$p$	$\neg p$
Do $\varphi$	-5	-5
Do $\neg\varphi$	-5	-5
Find out which is harmful	-1	-1

In fact, the doctor is making a decision under certainty. She knows that the value of prescribing either medicine is -5, and the value of running the tests is -1, so she should run the tests.

In general, when an agent has a duty to maximise the expected value of some quantity  $Q$ , then the value that goes into the agent's decision table in a cell is *not* the value of  $Q$  in the world-action pair the agent represents. Rather, it's the expected value of  $Q$  given that world-action pair. In situations like this one where the relevant facts (e.g., which medicine the patient is allergic to) don't affect the evidence the agent has, the decision is a decision under *certainty*. This is all as things should be. When you have obligations that are drawn in terms of the expected value of a variable, the actual values of that variable cease to be directly relevant to the decision problem.

One upshot of these considerations is that when moral and epistemic considerations get entangled, for example when agents have a moral duty not to take certain kinds of risks, it can get tricky to apply the theory of belief developed here. In a separate paper ("Defending Interest-Relative Invariantism") I've shown how this idea can help respond to some criticisms of similar views raised by Jessica Brown (2008).

### 3 Fantl and McGrath on Interest-Relativity

Jeremy Fantl and Matthew McGrath (2009) have argued that my interest-relative theory of belief cannot explain all of the interest-relativity in epistemology. I'm going to agree with their conclusion, but not with their premises. At this point you might suspect, dear reader, that you're about to be drawn into a micro-battle between two similar but not quite identical explanations of the same (alleged) phenomena. And you wouldn't be entirely mistaken. But I think seeing why Fantl and McGrath's objections to my theory fail will show us something interesting about the relationship between interests and knowledge. In particular, it will show us that interests can generate certain kinds of *defeaters* of claims to knowledge.

Fantl and McGrath's primary complaint against the interest-relative theory of belief I developed in my Weatherston (2005) and in the previous section is that it is not strong enough to entail principles such as (JJ).

(JJ) If you are justified in believing that  $p$ , then  $p$  is warranted enough to justify you in  $\varphi$ -ing, for any  $\varphi$ . (Fantl and McGrath, 2009, 99)

In practice, what this means is that there can't be a salient  $p, \varphi$  such that:

- The agent is justified in believing  $p$ ;
- The agent is not warranted in doing  $\varphi$ ; but
- If the agent had more evidence for  $p$ , and nothing else, the agent would be warranted in doing  $\varphi$ .

That is, once you've got enough evidence, or warrant, for justified belief in  $p$ , then you've got enough evidence for  $p$  as matters for any decision you face. This seems intuitive, and Fantl and McGrath back up its intuitiveness with some nicely drawn examples.

Now it's true that the interest-relative theory of belief cannot be used to derive (JJ), at least on the reading of it I just provided. But that's because on the intended reading, it is false, and the interest-relative theory is true. So the fact that (JJ) can't be derived is a feature, not a bug. The problem arises because of cases like that of Coraline. Here's what we're going to stipulate about Coraline.

- She knows that  $p$  and  $q$  are independent, so her credence in any conjunction where one conjunct is a member of  $\{p, \neg p\}$  and the other is a member of  $\{q, \neg q\}$  will be the product of her credences in the conjuncts.
- Her credence in  $p$  is 0.99, just as the evidence supports.
- Her credence in  $q$  is also 0.99. This is unfortunate, since the rational credence in  $q$  given her evidence is 0.01.
- She has a choice between taking and declining a bet with the following payoff structure.<sup>20</sup> (Assume that the marginal utility of money is close enough to constant that expected dollar returns correlate more or less precisely with expected utility returns.)

	$p \wedge q$	$p \wedge \neg q$	$\neg p$
<b>Take bet</b>	\$100	\$1	-\$1000
<b>Decline bet</b>	0	0	0

As can be easily computed, the expected utility of taking the bet given her credences is positive, it is just over \$89. And Coraline takes the bet. She doesn't compute the expected utility, but she is sensitive to it.<sup>21</sup> That is, had the expected utility given her credences been close to 0, she would have not acted until she made a computation. But from her perspective this looks like basically a free \$100, so she takes it. Happily, this all turns out well enough, since  $p$  is true. But it was a dumb thing to do. The expected utility of taking the bet given her evidence is negative, it is a little under -\$8. So she isn't warranted, given her evidence, in taking the bet.

I also claim the following three things are true of her.

<sup>20</sup>I'm more interested in the abstract structure of the case than in whether any real-life situation is modelled by just this structure. But it might be worth noting the rough kind of situation where this kind of situation can arise. So let's say Coraline has a particular bank account that is uninsured, but which currently paying 10% interest, and she is deciding whether to deposit another \$1000 in it. Then  $p$  is the proposition that the bank will not collapse, and she'll get her money back, and  $q$  is the proposition that the interest will stay at 10%. To make the model exact, we have to also assume that if the interest rate on her account doesn't stay at 10%, it falls to 0.1%. And we have to assume that the interest rate and the bank's collapse are probabilistically independent. Neither of these are at all realistic, but a realistic case would simply be more complicated, and the complications would obscure the philosophically interesting point.

<sup>21</sup>If she did compute the expected utility, then one of the things that would be salient for her is the expected utility of the bet. And the expected utility of the bet is different to its expected utility given  $p$ . So if that expected utility is salient, she doesn't believe  $p$ . And it's going to be important to what follows that she *does* believe  $p$ .

1.  $p$  is not justified enough to warrant her in taking the bet.
2. She believes  $p$ .<sup>22</sup>
3. This belief is rational.

The argument for 1 is straightforward. She isn't warranted in taking the bet, so  $p$  isn't sufficiently warranted to justify it. This is despite the fact that  $p$  is obviously relevant. Indeed, given  $p$ , taking the bet strictly dominates declining it. But still,  $p$  doesn't warrant taking this bet, because nothing warrants taking a bet with negative expected utility. Had the rational credence in  $p$  been higher, then the bet would have been reasonable. Had the reasonable credence in  $p$  been, say, 0.9999, then she would have been reasonable in taking the bet, and using  $p$  as a reason to do so. So there's a good sense in which  $p$  simply isn't warranted enough to justify taking the bet.<sup>23</sup>

The argument for 2 is that she has a very high credence in  $p$ , this credence is grounded in the evidence in the right way, and it leads her to act as if  $p$  is true, e.g. by taking the bet. It's true that her credence in  $p$  is not 1, and if you think credence 1 is needed for belief, then you won't like this example. But if you think that, you won't think there's much connection between (JJ) and pragmatic conditions in epistemology either. So that's hardly a position a defender of Fantl and McGrath's position can hold.<sup>24</sup>

The argument for 3 is that her attitude towards  $p$  tracks the evidence perfectly. She is making no mistakes with respect to  $p$ . She is making a mistake with respect to  $q$ , but not with respect to  $p$ . So her attitude towards  $p$ , i.e. belief, is rational.

I don't think the argument here strictly needs the assumption I'm about to make, but I think it's helpful to see one very clear way to support the argument of the last paragraph. The working assumption of my project on interest-relativity has been that talking about beliefs and talking about credences are simply two ways of modelling the very same things, namely minds. If the agent both has a credence 0.99 in  $p$ , and believes that  $p$ , these are not two different states. Rather, there is one state of the agent, and two different ways of modelling it. So it is implausible, if not incoherent, to apply different valuations to the state depending on which modelling tools we choose to use. That is, it's implausible to say that while we're modelling the agent with credences, the state is rational, but when we change tools, and start using beliefs, the state is irrational. Given this outlook on beliefs and credences, premise 3 seems to follow immediately from the setup of the example.

So that's the argument that (JJ) is false. And if it's false, the fact that the interest-relative theory doesn't entail it is a feature, not a bug. But there are a number of

<sup>22</sup>In terms of the example discussed in the previous footnote, she believes that the bank will survive, i.e., that she'll get her money back if she deposits it.

<sup>23</sup>I'm assuming here that the interpretation of (JJ) that I gave above is correct, though actually I'm not entirely sure about this. For present purposes, I plan to simply interpret (JJ) that way, and not return to exegetical issues.

<sup>24</sup>We do have to assume that  $\neg q$  is not so salient that attitudes conditional on  $\neg q$  are relevant to determining whether she believes  $p$ . That's because conditional on  $\neg q$ , she prefers to not take the bet, but conditional on  $\neg q \wedge p$ , she prefers to take the bet. But if she is simply looking at this as a free \$100, then it's plausible that  $\neg q$  is not salient.

possible objections to that position. I'll spend the rest of this section, and this paper, going over them.<sup>25</sup>

*Objection:* The following argument shows that Coraline is not in fact justified in believing that  $p$ .

1.  $p$  entails that Coraline should take the bet, and Coraline knows this.
  2. If  $p$  entails something, and Coraline knows this, and she justifiably believes  $p$ , she is in a position to justifiably believe the thing entailed.
  3. Coraline is not in a position to justifiably believe that she should take the bet.
- C. So, Coraline does not justifiably believe that  $p$

*Reply:* The problem here is that premise 1 is false. What's true is that  $p$  entails that Coraline will be better off taking the bet than declining it. But it doesn't follow that she should take the bet. Indeed, it isn't actually true that she should take the bet, even though  $p$  is actually true. Not just is the entailment claim false, the world of the example is a counterinstance to it.

It might be controversial to use this very case to reject premise 1. But the falsity of premise 1 should be clear on independent grounds. What  $p$  entails is that Coraline will be best off by taking the bet. But there are lots of things that will make me better off that I shouldn't do. Imagine I'm standing by a roulette wheel, and the thing that will make me best off is betting heavily on the number that will actually come up. It doesn't follow that I should do that. Indeed, I should not do it. I shouldn't place any bets at all, since all the bets have a highly negative expected return.

In short, all  $p$  entails is that taking the bet will have the best consequences. Only a very crude kind of consequentialism would identify what I should do with what will have the best returns, and that crude consequentialism isn't true. So  $p$  doesn't entail that Coraline should take the bet. So premise 1 is false.

*Objection:* Even though  $p$  doesn't entail that Coraline should take the bet, it does provide inductive support for her taking the bet. So if she could justifiably believe  $p$ , she could justifiably (but non-deductively) infer that she should take the bet. Since she can't justifiably infer that, she isn't justified in taking the bet.

*Reply:* The inductive inference here looks weak. One way to make the inductive inference work would be to deduce from  $p$  that taking the bet will have the best outcomes, and infer from that that the bet should be taken. But the last step doesn't even look like a reliable ampliative inference. The usual situation is that the best outcome comes from taking an *ex ante* unjustifiable risk.

It may seem better to use  $p$  combined with the fact that conditional on  $p$ , taking the bet has the highest *expected* utility. But actually that's still not much of a reason to take the bet. Think again about cases, completely normal cases, where the action

<sup>25</sup>Thanks here to a long blog comments thread with Jeremy Fantl and Matthew McGrath for making me formulate these points much more carefully. The original thread is at <http://tar.weatherson.org/2010/03/31/do-justified-beliefs-justify-action/>.

with the best outcome is an *ex ante* unjustifiable risk. Call that action  $\varphi$ , and let  $B\varphi$  be the proposition that  $\varphi$  has the best outcome. Then  $B\varphi$  is true, and conditional on  $B\varphi$ ,  $\varphi$  has an excellent expected return. But doing  $\varphi$  is still running a dumb risk. Since these kinds of cases are normal, it seems it will very often be the case that this form of inference leads from truth to falsity. So it's not a reliable inductive inference.

More generally, we should worry quite a lot about Coraline's ability to draw inductive inferences about the propriety of the bet here. Unlike deductive inferences, inductive inferences can be defeated by a whole host of factors. If I've seen a lot of swans, in a lot of circumstances, and they've all been blue, that's a good reason to think the next swan I see will be blue. But it ceases to be a reason if I am told by a clearly reliable testifier that there are green swans in the river outside my apartment. And that's true even if I dismiss the testifier because I think he has a funny name, and I don't trust people with funny names. Now although Coraline has evidence for  $p$ , she also has a lot of evidence against  $q$ , evidence that she is presumably ignoring since her credence in  $q$  is so high. Any story about how Coraline can reason from  $p$  to the claim that she should have to take the bet will have to explain how her irrational attraction to  $q$  doesn't serve as a defeater, and I don't see how that could be done.

*Objection:* In the example, Coraline isn't just in a position to justifiably believe  $p$ , she is in a position to *know* that she justifiably believes it. And from the fact that she justifiably believes  $p$ , and the fact that if  $p$ , then taking the bet has the best option, she can infer that she should take the bet.

*Reply:* It's possible at this point that we get to a dialectical impasse. I think this inference is non-deductive, because I think the example we're discussing here is one where the premises are true and the conclusion false. Presumably someone who doesn't like the example will think that it is a good deductive inference.

What makes the objection useful is that, unlike the inductive inference mentioned in the previous objection, this at least has the *form* of a good inductive inference. Whenever you justifiably believe  $p$ , and the best outcome given  $p$  is gained by doing  $\varphi$ , then *usually* you should  $\varphi$ . Since Coraline knows the premises are true, *ceteris paribus* that gives her a reason to believe the premise is probably true.

But other things aren't at all equal. In particular, this is a case where Coraline has a highly irrational credence concerning a proposition whose probability is highly relevant to the expected utility of possible actions. Or, to put things another way, an inference from something to something else it is correlated with can be defeated by related irrational beliefs. (That's what the swan example above shows.) So if Coraline tried to infer this way that she should take the bet, her irrational confidence in  $q$  would defeat the inference.

The objector might think I am being uncharitable here. The objection doesn't say that Coraline's knowledge provides an *inductive* reason to take the bet. Rather, they say, it provides a *conclusive* reason to take the bet. And conclusive reasons cannot be defeated by irrational beliefs elsewhere in the web. Here we reach an impasse. I say that knowledge that you justifiably believe  $p$  cannot provide a conclusive reason to bet on  $p$  because I think Coraline knows she justifiably believes  $p$ , but does not

have a conclusive reason to bet on  $p$ . That if, I think the premise the objector uses here is false because I think (JJ) is false. The person who believes in (JJ) won't be so impressed by this move.

Having said all that, the more complicated example at the end of Weatherson (2005) was designed to raise the same problem without the consequence that if  $p$  is true, the bet is sure to return a positive amount. In that example, conditionalising on  $p$  means the bet has a positive expected return, but still possibly a negative return. But in that case (JJ) still failed. If accepting there are cases where an agent justifiably believes  $p$ , and knows this, but can't rationally bet on  $p$  is too much to accept, that more complicated example might be more persuasive. Otherwise, I concede that someone who believes (JJ) and thinks rational agents can use it in their reasoning will not think that a particular case is a counterexample to (JJ).

*Objection:* If Coraline were ideal, then she wouldn't believe  $p$ . That's because if she were ideal, she would have a lower credence in  $q$ , and if that were the case, her credence in  $p$  would have to be much higher (close to 0.999) in order to count as a belief. So her belief is not justified.

*Reply:* The premise here, that if Coraline were ideal she would not believe that  $p$ , is true. The conclusion, that she is not justified in believing  $p$ , does not follow. It's always a mistake to *identify* what should be done with what is done in ideal circumstances. This is something that has long been known in economics. The *locus classicus* of the view that this is a mistake is Lipsey and Lancaster (1956-1957). A similar point has been made in ethics in papers such as Watson (1977) and Kennett and Smith (1996a,b). And it has been extended to epistemology by Williamson (1998).

All of these discussions have a common structure. It is first observed that the ideal is both  $F$  and  $G$ . It is then stipulated that whatever happens, the thing being created (either a social system, an action, or a cognitive state) will not be  $F$ . It is then argued that given the stipulation, the thing being created should not be  $G$ . That is not just the claim that we shouldn't *aim* to make the thing be  $G$ . It is, rather, that in many cases being  $G$  is not the best way to be, given that  $F$ -ness will not be achieved. Lipsey and Lancaster argue that (in an admittedly idealised model) that it is actually quite unusual for  $G$  to be best given that the system being created will not be  $F$ .

It's not too hard to come up with examples that fit this structure. Following (Williamson, 2000, 209), we might note that I'm justified in believing that there are no ideal cognitive agents, although were I ideal I would not believe this. Or imagine a student taking a ten question mathematics exam who has no idea how to answer the last question. She knows an ideal student would correctly answer an even number of questions, but that's no reason for her to throw out her good answer to question nine. In general, once we have stipulated one departure from the ideal, there's no reason to assign any positive status to other similarities to the idea. In particular, given that Coraline has an irrational view towards  $q$ , she won't perfectly match up with the ideal, so there's no reason it's good to agree with the ideal in other respects, such as not believing  $p$ .

Stepping back a bit, there's a reason the interest-relative theory says that the ideal and justification come apart right here. On the interest-relative theory, like on any pragmatic theory of mental states, the *identification* of mental states is a somewhat holistic matter. Something is a belief in virtue of its position in a much broader network. But the *evaluation* of belief is (relatively) atomistic. That's why Coraline is justified in believing  $p$ , although if she were wiser she would not believe it. If she were wiser, i.e., if she had the right attitude towards  $q$ , the very same credence in  $p$  would not count as a belief. Whether her state counts as a belief, that is, depends on wide-ranging features of her cognitive system. But whether the state is justified depends on more local factors, and in local respects she is doing everything right.

*Objection:* Since the ideal agent in Coraline's position would not believe  $p$ , it follows that there is no *propositional* justification for  $p$ . Moreover, doxastic justification requires propositional justification<sup>26</sup> So Coraline is not doxastically justified in believing  $p$ . That is, she isn't justified in believing  $p$ .

*Reply:* I think there are two ways of understanding 'propositional justification'. On one of them, the first sentence of the objection is false. On the other, the second sentence is false. Neither way does the objection go through.

The first way is to say that  $p$  is propositionally justified for an agent iff that agent's evidence justifies a credence in  $p$  that is high enough to count as a belief *given the agent's other credences and preferences*. On that understanding,  $p$  is propositionally justified by Coraline's evidence. For all that evidence has to do to make  $p$  justified is to support a credence a little greater than 0.9. And by hypothesis, the evidence does that.

The other way is to say that  $p$  is propositionally justified for an agent iff that agent's evidence justifies a credence in  $p$  that is high enough to count as a belief *given the agent's preferences and the credences supported by that evidence*. On this reading, the objection reduces to the previous objection. That is, the objection basically says that  $p$  is propositionally justified for an agent iff the ideal agent in her situation would believe it. And we've already argued that that is compatible with doxastic justification. So either the objection rests on a false premise, or it has already been taken care of.

*Objection:* If Coraline is justified in believing  $p$ , then Coraline can use  $p$  as a premise in practical reasoning. If Coraline can use  $p$  as a premise in practical reasoning, and  $p$  is true, and her belief in  $p$  is not Gettiered, then she knows  $p$ . By hypothesis, her belief is true, and her belief is not Gettiered. So she should know  $p$ . But she doesn't know  $p$ . So by several steps of modus tollens, she isn't justified in believing  $p$ .<sup>27</sup>

*Reply:* Like the previous objection, this one turns on an equivocation, this time over the neologism 'Gettiered'. Some epistemologists use this to simply mean that a belief is justified and true without constituting knowledge. By that standard, the third sentence is false. Or, at least, we haven't been given any reason to think that it is true.

<sup>26</sup>See Turri (2010) for a discussion of recent views on the relationship between propositional and doxastic justification. This requirement seems to be presupposed throughout that literature.

<sup>27</sup>Compare the 'subtraction argument' on page 99 of Fantl and McGrath (2009).

Given everything else that's said, the third sentence is a raw assertion that Coraline knows that  $p$ , and I don't think we should accept that.

The other way epistemologists sometimes use the term is to pick out justified true beliefs that fail to be knowledge for the reasons that the beliefs in the original examples from Gettier (1963) fail to be knowledge. That is, it picks out a property that beliefs have when they are derived from a false lemma, or whatever similar property is held to be doing the work in the original Gettier examples. Now on this reading, Coraline's belief that  $p$  is not Gettiered. But it doesn't follow that it is known. There's no reason, once we've given up on the JTB theory of knowledge, to think that whatever goes wrong in Gettier's examples is the *only* way for a justified true belief to fall short of knowledge. It could be that there's a practical defeater, as in this case. So the second sentence of the objection is false, and the objection again fails.

But note I'm conceding to the objector that Coraline does not know  $p$ . That's important, because it reveals another way in which knowledge is interest-relative. We can see that Coraline does not know that  $p$  by noting that if she did know  $p$ , we could represent her decision like this:

	$q$	$\neg q$
Take bet	\$100	\$1
Decline bet	0	0

And now she should clearly take the bet, since it is a dominating option. Her irrational credence in  $q$  simply wouldn't matter. But she can't rationally take the bet, so this table must be wrong, so she doesn't know  $p$ .

This seems odd. Why should an irrational credence in  $q$  defeat knowledge of  $p$ ? The answer is that beliefs have to sufficiently cohere with our other beliefs to be knowledge. If I have a very firm belief that  $\neg p$ , and also a belief that  $p$ , the latter belief cannot be knowledge even if it is grounded in the facts in just the right way. The contradictory belief that  $\neg p$  is a doxastic defeater.

Now in practice almost all of our beliefs are in some tension with some other subset of our beliefs. Unless we are perfectly coherent, which we never are, there will be lots of ways to argue that any particular belief does not sufficiently cohere with our broader doxastic state, and hence is defeated from being knowledge. There must be some restrictions on when incoherence with some part of the rest of one's beliefs defeats knowledge.

I think an interest-relative story here is most likely to succeed. If we said Coraline knew that  $p$ , that would produce a tension between the fact that she does (rationally) believe that taking the bet is best given  $p$ , and that she should believe that declining the bet is not best *simpliciter*. And that matters because whether to take the bet or not is a live decision for her. That is, incoherencies or irrationalities that manifest in decision problems that are salient can defeat knowledge.



Put another way, an agent only knows  $p$  if we can properly model any decision she's really facing with a decision table where  $p$  is taken as given.<sup>28</sup> Crediting Cora-line with knowing  $p$  will mean we get the table for this decision, i.e., whether to take this very bet, wrong. And that's a live decision for her. Those two facts together entail that she doesn't know  $p$ . Neither alone would suffice for defeating knowledge of  $p$ . And that's a part of the explanation for the interest-relativity of knowledge, a part I left out of the story in my (2005).

---

<sup>28</sup>Or, at least, we can model any such decision where standard decision theory applies. Perhaps we'll have to exclude cases where one of the options violates a deontological constraint. As noted in section 1, the argument for interest-relativity of knowledge is neutral on how to handle such cases.

# Bibliography

- Blome-Tillmann, Michael. 2009. "Contextualism, Subject-Sensitive Invariantism, and the Interaction of 'Knowledge'-Ascriptions with Modal and Temporal Operators." *Philosophy and Phenomenological Research* 79:315–331, doi:10.1111/j.1933-1592.2009.00280.x.
- Braddon-Mitchell, David and Jackson, Frank. 2007. *The Philosophy of Mind and Cognition, second edition*. Malden, MA: Blackwell.
- Brown, Jessica. 2008. "Knowledge and Practical Reason." *Philosophy Compass* 3:1135–1152, doi:10.1111/j.1747-9991.2008.00176.x.
- Cohen, Stewart. 1988. "How to be a Fallibilist." *Philosophical Perspectives* 2:91–123.
- DeRose, Keith. 1995. "Solving the Skeptical Problem." *Philosophical Review* 104:1–52.
- Fantl, Jeremy and McGrath, Matthew. 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Feltz, Adam and Zarpentine, Chris. forthcoming. "Do You Know More When It Matters Less?" *Philosophical Psychology* Retrieved from <http://faculty.schreiner.edu/adfeltz/Papers/Know%20more.pdf>.
- Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23:121–123, doi:10.2307/3326922.
- Hammond, Peter J. 1988. "Consequentialist Foundations for Expected Utility." *Theory and Decision* 25:25–78, doi:10.1007/BF00129168.
- Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hawthorne, John and Stanley, Jason. 2008. "Knowledge and Action." *Journal of Philosophy* 105:571–90.
- Jackson, Frank. 1991. "Decision Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101:461–82.

- Kennett, Jeanette and Smith, Michael. 1996a. "Frog and Toad Lose Control." *Analysis* 56:63–73, doi:10.1111/j.0003-2638.1996.00063.x.
- . 1996b. "Philosophy and Commonsense: The Case of Weakness of Will." In Michaelis Michael and John O'Leary-Hawthorne (eds.), *The Place of Philosophy in the Study of Mind*, 141–157. Norwell, MA: Kluwer, doi:10.1017/CBO9780511606977.005.
- Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.
- . 1982. "Logic for Equivocators." *Noûs* 16:431–441. Reprinted in *Papers in Philosophical Logic*, pp. 97–110.
- . 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74:549–567, doi:10.1080/00048409612347521. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 418–446.
- Lipsey, R. G. and Lancaster, Kelvin. 1956–1957. "The General Theory of Second Best." *Review of Economic Studies* 24:11–32, doi:10.2307/2296233.
- Maher, Patrick. 1993. *Betting on Theories*. Cambridge: Cambridge University Press.
- Ramsey, Frank. 1926. "Truth and Probability." In D. H. Mellor (ed.), *Philosophical Papers*, 52–94. Cambridge: Cambridge University Press.
- Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.
- Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford University Press.
- Turri, John. 2010. "On the Relationship between Propositional and Doxastic Justification." *Philosophy and Phenomenological Research* 80:312–326, doi:10.1111/j.1933-1592.2010.00331.x.
- Watson, Gary. 1977. "Skepticism about Weakness of Will." *Philosophical Review* 86:316–339, doi:10.2307/2183785.
- Weatherson, Brian. 2003. "What Good Are Counterexamples?" *Philosophical Studies* 115:1–31, doi:10.1023/A:1024961917413.
- . 2005. "Can We Do Without Pragmatic Encroachment?" *Philosophical Perspectives* 19:417–443, doi:10.1111/j.1520-8583.2005.00068.x.
- . 2006. "Questioning Contextualism." In Stephen Cade Hetherington (ed.), *Epistemology Futures*, 133–147. Oxford: Oxford University Press.
- Williamson, Timothy. 1998. "Conditionalizing on Knowledge." *British Journal for the Philosophy of Science* 49:89–121, doi:10.1093/bjps/49.1.89.
- . 2000. *Knowledge and its Limits*. Oxford University Press.