Knowledge, Bets and Interests

Brian Weatherson

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.[1]

(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 (K) is context sensitive,

[1]The reflections in the next few paragraphs are inspired by some comments by Stalnaker in his (2008), though I don't want to suggest the theory I'll discuss is actually Stalnaker's.

and not merely 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**. Note that the way we're classifying theories, a view that holds that (K) is context-sensitive both because (L) is context-sensitive *and* because Cohen *et al* are correct is a version of non-doxastic contextualism. The label 'non-doxastic' is being used to mean that the contextualism isn't solely doxastic, rather than as denying contextualism about belief ascriptions.

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.

Again, a theory that holds both that belief ascriptions are interest-relative, and that some of the interest-relativity of knowledge ascriptions is not explained by the interest-relativity of belief ascriptions, will count as a version of non-doxastic IRI. I'm going to defend a view from this class here.

In my Weatherson (2005) I tried to motivate Doxastic IRI. It isn't completely trivial to map my view onto the existing views in the literature, but the idea was to renounce contextualism and all its empty promises, and endorse a position that's usually known as 'strict invariantism' about these classes of statements:

- *S* is justified in having credence *x* in *p*;
- If *S* believes that *p*, she knows that *p*;

while holding that the interests of S are relevant to the truth of statements from these classes:

- *S* believes that *p*;
- *S* justifiably believes that *p*;
- *S* knows that *p*.

But I didn't argue for all of that. What I argued for was Doxastic IRI about ascriptions of justified belief, and I 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.[2] My change of heart has been

---

[2]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. Whether I can hold onto the strict invariantism about claims about justified credences depends, I now think, on whether an interest-neutral account of evidence can be given. Discussions with Tom Donaldson and Jason Stanley have left me less convinced than I was in 2005 that this is possible, but this is far too big a question to resolve here.

prompted by cases like those Jason Stanley (2005) calls 'Ignorant High Stakes' cases.[3] But to see why these cases matter, it will help to start with why I think some kind of IRI must be true.

Here's the plan of attack. In section 1, I'm going to argue that knowledge plays an important role in decision theory. In particular, I'll argue (a) that it is legitimate to write something onto a decision table iff the decision maker knows it to be true, and (b) it is legitimate to leave a possible state of the world off a decision table iff the decision maker knows it not to obtain. I'll go on to argue that this, plus some very plausible extra assumptions about the rationality of certain possible choices, implies that knowledge is interest-relative. In section 2 I'll summarise and extend the argument from Weatherson (2005) that belief is interest-relative. People who are especially interested in the epistemology rather than the theory of belief may skip this. But I think this material is important; most of the examples of interest-relative knowledge in the literature can be explained by the interest-relativity of belief. I used to think all such cases could be explained. Section 3 describes why I no longer think that. Reflections on cases like the Coraline example suggests that there are coherence constraints on knowledge that go beyond the coherence constraints on justified true belief. The scope of these constraints is, I'll argue, interest-relative. So knowledge, unlike belief or justified belief, has interest-relative defeaters. That's inconsistent with Doxastic IRI, so Doxastic IRI is false.

1 The Interest-Relativity of Knowledge

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

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.

---

[3]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.

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 the agent won't 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.

But Stu is not the only problem that Professor Dec has. She also has trouble convincing Dom of the argument. He thinks there should be a third state added, S_3. In S_3, there is a vengeful God who is about to end the world, and take everyone who chose C_1 to heaven, while sending everyone who chose C_2 to hell. Since heaven is better than hell, C_2 does not dominate C_1; it is worse in S_3. If decision theory is to be useful, we must say something about why we can leave states like S_3 off the decision table.

So in order to teach decision theory, Professor Dec has to answer two questions.[4]

1. What makes it legitimate to write something on the decision table, such as the '$1500' we write in the bottom right cell of Dec's table?
2. What makes it legitimate to leave something off a decision table, such as leaving Dom's state S_3 off the table?

Let's start with a simpler problem that helps with both questions. 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

---

[4]If we are convinced that the right decision is the one that maximises expected utility, there is a sense in which these questions collapse. For the expected utility theorist, we can solve Dom's question by making sure the states are logically exhaustive, and making the 'payouts' in each state be expected payouts. But the theory that the correct decision is the one that maximises expected utility, while plausibly true, is controversial. It shouldn't be assumed when we are investigating the semantics of decision tables.

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 28th, 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.[5] 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 is 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 rational 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 her students 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. And the miscommunication between Professor Dec and Dom also concerns knowledge. When Dec wrote that table up, she was saying that the agent knew that S_1 or S_2 obtained. And when she says it is best to take dominating options, she means that it is best to take options that one knows to have better outcomes. So here are the answers to Stu and Dom's challenges.

1. It is legitimate to write something on the decision table, such as the '$1500' we write in the bottom right cell of Dec's table, iff the decision maker knows it to be true.
2. It is legitimate to leave something off a decision table, such as leaving Dom's state S_3 off the table, iff the decision maker knows it not to obtain.

---

[5]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.

Perhaps those answers are not correct, but what we can clearly see by reflecting on these cases is that 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 the information explicitly on the table (such as that Alice will get \$1500 if C_2 and S_2) and implied by where the table ends (such as that S_3 will not happen). 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.[6]

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

---

[6]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 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.

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 this footnote 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.

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 making a particular decision when they stand in these very strong relationships to the parameters of the decision table, then people will almost never be making the kinds of decision we study in decision theory. Since decision theory and game 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 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? The response I've offered to Dom's challenge suggests 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.

We might be tempted to say that we can abstract away from anything such that the difference between its probability and 1 doesn't make a difference to the ultimate answer to the decision problem. More carefully, the idea would be that we can have the decision table represent that p iff p is true and treating $\Pr(p)$ as 1 rather than its actual value doesn't change what the agent should do. I think this is the most plausible story one could tell about decision tables if one didn't like the knowledge first story that I tell. But I also don't think it works, because of cases like the following.

Luc is lucky; he's in a casino where they are offering better than fair odds on roulette. Although the chance of winning any bet is ⅟₃₈, if Luc bets $10, and his bet

wins, he will win \$400. (That's the only bet on offer.) Luc, like Alice, is considering betting on 28. As it turns out, 28 won't come up, although since this is a fair roulette wheel, Luc doesn't know this. Luc, like most agents, has a declining marginal utility for money. He currently has \$1,000, and for any amount of money x, Luc gets utility $u(x) = x^{1/2}$ out of having x. So Luc's current utility (from money) is, roughly, 31.622. If he bets and loses, his utility will be, roughly, 31.464. And if he bets and wins, his utility will be, roughly, 37.417. So he stands to gain about 5.794, and to lose about 0.159. So he stands to gain about 36.5 as much as he stands to lose. Since the odds of winning are less than $\frac{1}{36.5}$, his expected utility goes down if he takes the bet, so he shouldn't take it. Of course, if the probability of losing was 1, and not merely $\frac{37}{38}$, he shouldn't take the bet too. Does that mean it is acceptable, in presenting Luc's decision problem, to leave off the table any possibility of him winning, since he won't win, and setting the probability of losing to 1 rather than $\frac{37}{38}$ doesn't change the decision he should make? Of course not; that would horribly misstate the situation Luc finds himself in. It would misrepresent how sensitive Luc's choice is to his utility function, and to the size of the stakes. If Luc's utility function was $u(x) = x^{3/4}$, then he should take the bet. If his utility function is unchanged, but the bet was \$1 against \$40, rather than \$10 against \$400, he should take the bet. Leaving off the possibility of winning hides these facts, and badly misrepresents Luc's situation.

I've argued that the states we can 'leave off' a decision table are the states that the agent knows not to obtain. The argument is largely by elimination. If we can only leave off things that have probability 1, then decision theory would be useless; but it isn't. If we say we can leave off things if setting their probability at 1 is an accepable idealisation, we need a theory of acceptable idealisations. If this is to be a rival to my theory, the idealisation had better not be it's acceptable to treat anything known as having probability 1. But the most natural alternative idealisation badly misrepresents Luc's case. If we say that what can be left off is not what's known not to obtain, but what is, say, justifiably truly believed not to obtain, we need an argument for why people would naturally use such an unnatural standard. This doesn't even purport to be a conclusive argument, but these considerations point me towards thinking that knowledge determines what we can leave off.

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 incorporated into the table we use to represent her decision. 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.[7] To focus intuitions, let's take a concrete case. Barry is sitting in his apartment one evening when he hears

---

[7]This issue is of course central to the plotline in Hawthorne (2004).

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 an offers him a bet.[8] If he takes the bet, and the musician is Beth, then the genie will give 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.

	Musician is Beth	Musician is not Beth
Take Bet	Win $10	1000 years of torture
Decline Bet	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. So let's write the table in the standard form.

	Musician is Beth
Take Bet	Win $10
Decline Bet	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 it seems to me that there are three options available.

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.

---

[8]Assume, perhaps implausibly, that the sudden appearance of the genie is evidentially irrelevant to the proposition that the musician is Beth. The reasons this may be implausible are related to the arguments in (Runyon, 1992, 14-15). Thanks here to Jeremy Fantl.

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.[9] And the third option would send us back into the problems about interpreting and applying decision theory that we spent the first few pages trying to get out of.

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. When Barry was unconcerned with bets at extremely long odds on whether the musician is Beth, he knows Beth is the musician. Now that he is interested in those bets, he doesn't know that.[10]

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 can get to interest-relativity with very weak premises.[11] 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'. That is, our decision tables should include what the agent knows about the payoffs in different states, and leave off any state the agent knows not to obtain.
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

---

[9]The idea that interest-relativity is a way of fending off scepticism is a very prominent theme in Fantl and McGrath (2009).

[10]On the version of IRI I'm defending, Barry is free to be interested in whatever he likes. If he started wondering about whether it would be rational to take such a bet, he loses the knowledge that Beth is the musician, even if there is no genie and the bet isn't offered. The existence of the genie's offer makes the bet a practical interest; merely wondering about the genie's offer makes the bet a cognitive interest. But both kinds of interests are relevant to knowledge.

[11]As they make clear in their (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. The strategy here isn't a million miles from the point noted in Fantl and McGrath (2009, 72n14) when they note that much weaker premises than the ones they endorse imply a failure of 'purism'.

anything about what else might or might not make acceptance rationally permissible. It doesn't say anything about what suffices for rationally 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.[12] 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.[13]

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 of interest-relativity about knowledge. That is, my view is a form of interest-relative-invariantism, or IRI.[14]

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.[15] 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.

---

[12]I have more to say about those cases in section 2.2.

[13]Also note that I'm not taking as a premise any claim about what Barry knows after the bet is offered. A lot of work on interest-relativity has used such premises, or premises about related intuitions. This seems like a misuse of the method of cases to me. That's not because we should never use intuitions about cases, just that these cases are too hard to think that snap judgments about them are particularly reliable. In general, we can know a lot about cases by quickly reflecting on them. Similarly, we know a lot about which shelves are level and which are uneven by visual inspection, i.e., 'eyeballing'. But when different eyeballs disagree, it's time to bring in other tools. That's the approach of this paper. I don't have a story about why the various eyeballs disagree about cases like Barry's; that seems like a task best undertaken by a psychologist not a philosopher (Ichikawa, 2009).

[14]This is obviously not a full argument against contextualism; that would require a much longer paper than this.

[15]See, for instance, Blome-Tillmann (2009), or Feltz and Zarpentine (2010).

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. It's the second part, the part that is new to this paper, that makes the theory a version of non-doxastic IRI.

Here's my theory of belief. *S* believes that *p* iff conditionalising on *p* doesn't change *S*'s answer to any relevant question. I'm using 'relevance' here in a non-technical sense; I say a lot more about how to cash out the notion in my (2005). The key thing to note is that relevance is interest-relative, so the theory of belief is interest-relative. There is a bit more to say about what kind of *questions* are important for this definition of belief. In part because I've changed my mind a little bit on this since the earlier paper, I'll spend a bit more time on it. The following four kinds of questions are the most important.

- How probable is q?
- Is q or r more probable?
- How good an idea is it to do ϕ?
- Is it better to do ϕ or ψ?

The theory of belief says that someone who believes that *p* doesn't change their answer to any of these questions upon conditionalising on *p*. Putting this formally, and making the restriction to relevant questions explicit, we get the following theorems of our theory of belief.[16]

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 $\Pr(q|p) \geq \Pr(r|p)$.

---

[16]In the last two lines, I use $U(\phi)$ to denote the expected utility of ϕ, and $U(\phi|p)$ to denote the expected utility of ϕ conditional on p. It's often easier to write this as simply $U(\phi \wedge p)$, since the utility of ϕ conditional on p just is the utility of doing ϕ 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 explictly conditional notation for each.

BAU For all relevant ϕ, x, if p is believed then $U(\phi) = x$ iff $U(\phi|p) = x$.
BCU For all relevant ϕ, ψ, if p is believed then $U(\phi) \geq U(\psi)$ iff $U(\phi|p) \geq U(\psi|p)$.

In the earlier paper I focussed on **BAU** and **BCU**. But **BAP** and **BCP** are important as well. Indeed, focussing on them lets us derive a nice result.

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. Indeed, in some circumstances she can rationally believe something *in virtue* of it being 0.978 probable.

That's because 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.[17] 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.

So there are some nice features of this theory of belief. Indeed, there are several reasons to believe it. It is, I have argued, the best functionalist account of belief. I'm not going to argue for functionalism about the mind, 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 do think functionalism is true, and so the best functionalist theory of belief is the best theory of belief.

The argument for this theory of belief in my (2005) rested heavily on the flaws of rival theories. We can see those flaws by looking at a tension that any theory of the relationship between belief and credence must overcome. Each of the following three principles seems to be plausible.

1. If S has a greater credence in p than in q, and she believes q, then she believes p as well; and if her credences in both p and q are rational, and her belief in q is rational, then so is her belief in p.
2. If S rationally believes p and rationally believes q, then it is open to her to rationally believe $p \wedge q$ without changing her credences.
3. S can rationally believe p while having credence of less than 1 in p.

---

[17]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.

But these three principles, together with some principles that are genuinely uncontroversial, entail an absurd result. By 3, there is some p such that $Cr(p) = x < 1$, and p is believed. (Cr is the function from any proposition to our agent's credence in that propositions.) Let S know that a particular fair lottery has l tickets, where $l > 1/_{1-x}$. The uncontroversial principle we'll use is that in such a case S's credence that any given ticket will lose should be $^{l-1}/_l$. Since $^{l-1}/_l > x$, it follows by 1 that S believes of each ticket that it will lose. Since her credences are rational, these beliefs are rational. By repeated applications of 2 then, the agent can rationally believe that each ticket will lose. But she rationally gives credence 0 to the proposition that each ticket will lose. So by 1 she can rationally believe any proposition in which her credence is greater than 0. This is absurd.[18]

I won't repeat all the gory details here, but one of the consequences of the discussion in Weatherson (2005) was that we could hold on to 3, and onto restricted versions of 1 and 2. In particular, if we restricted 1 and 2 to relevant propositions (in some sense) they became true, although the unrestricted version is false. A key part of the argument of the earlier paper was that this was a better option than the more commonly taken option of holding on to unrestricted versions of 1 and 3, at the cost of abandoning 2 even in clear cases. But one might wonder why I'm holding so tightly on to 3. After all, there is a functionalist argument that 3 is false.

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 p utils. A key functional role of belief is that if an agent believes p, and recognises that ϕ is the best thing to do given p, then she'll do ϕ. 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 this argument isn't quite right. If we spell out more carefully what the functional roles of credence and belief are, a loophole emerges in the argument that belief implies credence 1. The interest-relative theory of belief turns out to exploit 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.[19] No matter; as long as one agent is *disposed* to accept the offer, and the 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 situa-

---

[18]See Sturgeon (2008) for discussion of a similar puzzle for anyone trying to tell a unified story of belief and credence.

[19]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.

tions 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.

One of the crucial features of the theory of belief I'm defending is that what an agent believes is in general *insensitive* to such abtruse dispositions, although it is very sensitive to dispositions about practical matters. 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? The theory being defended here says no, and that seems plausible. 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.[20]

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

Those look a lot like the theorems of the theory of belief that we discussed above. But note that these claims are *unrestricted*, whereas in the theory of belief, we restricted attention to relevant actions, propositions, utilities and probabilities. That turns out to be the difference between belief and credence 1. Since that difference is interest-relative, belief is interest-relative.

I used to think that that was all the interest-relativity we needed in epistemology. Now I don't, for reasons that I'll go through in section three. (Readers who care more about the theory of knowledge than the theory of belief may want to skip ahead to that section.) But first I want to clean up some loose ends in the acount of belief.

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 Charlie, 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

---

[20]The presentation in this section, as in the earlier paper, assumes at least a weak form of consequentialism in the sense of Hammond (1988). 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.

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.[21] 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 geography of the American southwest.

Here's a way out of this problem in four relatively easy steps.[22] 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 p is more probable than $\neg p$ is a relevant question. Third, we infer from that that an agent only believes p if their credence in p is greater than their credence in $\neg p$, i.e., if their credence in p is greater than ½. Finally, we say that when the issue is whether the subject believes that p, the question of whether p is more probable than $\neg p$ is not only relevant 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 at the start of this section. 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

---

[21]I'm borrowing this example from Fred Dretske, who uses it to make some interesting points about dispositional belief.

[22]The recipe here is similar to that given in Weatherson (2005), but the motivation is streamlined. Thanks to Jacob Ross for helpful suggestions here.

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 φ or ψ. Failure to do either of these will lead to disaster, and is clearly unacceptable. Either φ or ψ 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 φ and ψ is harmful and which is not for a nominal cost. Right now, her credence that φ is the harmful one is, quite reasonably, ½. So the agent has three choices:

- Do φ;
- Do ψ; 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 propostion that φ 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 φ	-10	0
Do ψ	0	-10
Find out which is harmful	-1	-1

Given that $Pr(p)$ = ½, it's easy to compute that the expected value of doing either φ or ψ 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.[23]

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* ψ. So you might think that conditional on p, the agent should do ψ. That is, you might think the conditional expected value of doing ψ, 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.

---

[23]Some consequentialists say that what the agent should do depends on whether p is true. If p is true, she should do ψ, and if p is false she should do φ. As we'll see, I have reasons for thinking this is rather radically wrong.

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 φ and ψ 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 is 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 the incorporate these ideas into the formal decision theory. Even under the assumption that p, prescribing ψ 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.

	p	$\neg p$
Do φ	-5	-5
Do ψ	-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.

Similar morals carry across to theories that offer a smaller role to expected utility in determining moral value. In particular, it's often true that decisions where it is uncertain what course of action will produce the best outcome might still, in the morally salient sense, be decisions under certainty. That's because the uncertainty doesn't impact how we should weight the different possible outcomes, as in orthodox utility theory, but how we should value them. That's roughly what I think is going on in cases like this one, which Jessica Brown has argued are problematic for the epistemological theories John Hawthorne and Jason Stanley have recently been defending.[24]

> A student is spending the day shadowing a surgeon. In the morning he observes her in clinic examining patient A who has a diseased left kidney. The decision is taken to remove it that afternoon. Later, the student observes the surgeon in theatre where patient A is lying anaesthetised on the operating table. The operation hasn't started as the surgeon is consulting the patient's notes. The student is puzzled and asks one of the nurses what's going on:
>
> **Student**: I don't understand. Why is she looking at the patient's records? She was in clinic with the patient this morning. Doesn't she even know which kidney it is?
>
> **Nurse**: Of course, she knows which kidney it is. But, imagine what it would be like if she removed the wrong kidney. She shouldn't operate before checking the patient's records. (Brown, 2008, 1144-1145)

It is tempting, but for reasons I've been going through here mistaken, to represent the surgeon's choice as follows. Let **Left** mean the left kidney is diseased, and **Right** mean the right kidney is diseased.

	Left	Right
Remove left kidney	1	−1
Remove right kidney	−1	1
Check notes	$1 - \varepsilon$	$1 - \varepsilon$

Here ε is the trivial but non-zero cost of checking the chart. Given this table, we might reason that since the surgeon knows that she's in the left column, and removing the left kidney is the best option in that column, she should remove the left kidney rather than checking the notes.

But that reasoning assumes that the surgeon does not have any epistemic obligations over and above her duty to maximise expected utility. And that's very implausible. It's totally implausible on a non-consequentialist moral theory. A non-consequentialist may think that some people have just the same obligations that the

---

[24]The target here is not directly the interest-relativity of their theories, but more general principles about the role of knowledge in action and assertion. Since my theories are close enough, at least in consequences, to Hawthorne and Stanley's, it is important to note how my theory handles the case.

consequentialist says they have – legislators are frequently mentioned as an example – but surely they wouldn't think *surgeons* are in this category. And even a consequentialist who thinks that surgeons have special obligations in terms of their institutional role should think that the surgeon's obligations go above and beyond the obligation every agent has to maximise expected utility.

It's not clear exactly what the obligation the surgeon has. Perhaps it is an obligation to not just know which kidney to remove, but to know this on the basis of evidence she has obtained while in the operating theatre. Or perhaps it is an obligation to make her belief about which kidney to remove as sensitive as possible to various possible scenarios. Before she checked the chart, this counterfactual was false: *Had she misremembered which kidney was to be removed, she would have a true belief about which kidney was to be removed.* Checking the chart makes that counterfactual true, and so makes her belief that the left kidney is to be removed a little more sensitive to counterfactual possibilities.

However we spell out the obligation, it is plausible given what the nurse says that the surgeon has some such obligation. And it is plausible that the 'cost' of violating this obligation, call it δ is greater than the cost of checking the notes. So here is the decision table the surgeon faces.

	Left	Right
Remove left kidney	$1 - \delta$	$-1 - \delta$
Remove right kidney	$-1 - \delta$	$1 - \delta$
Check notes	$1 - \varepsilon$	$1 - \varepsilon$

And it isn't surprising, or a problem for an interest-relative theory of knowledge or belief, that the surgeon should check the notes, even if she believes *and knows* that the left kidney is the diseased one.

3 Interest-Relative Defeaters

As I said at the top, I've changed my view from Doxastic IRI to Non-Doxastic IRI. The change of heart is occasioned by cases like the following, where the agent is mistaken, and hence ignorant, about the odds at which she is offered a bet on *p*. In fact the odds are much longer than she thinks. Relative to what she stands to win, the stakes are too high.

3.1 *The Coraline Example*

The problem for Doxastic IRI 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.

- The only relevant question for her which is sensitive to p is whether to take or decline a bet with the following payoff structure.[25] (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.)

	$p \land q$	$p \land \neg q$	$\neg p$
Take bet	100	1	1000
Decline bet	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.[26] 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.

3.2 What Coraline Knows and What She Believes

Assume, for *reductio*, that Coraline knows that p. Then the choice she faces looks like this.

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

Since taking the bet dominates declining the bet, she should take the bet if this is the correct representation of her situation. She shouldn't take the bet, so by *modus tollens*, that can't be the correct representation of her situation. If she knew p, that would be the correct representation of her situation. So, again by *modus tollens*, she doesn't know p.

Now let's consider three possible explanations of why she doesn't know that p.

1. She doesn't have enough evidence to know that p, independent of the practical stakes.

---

[25]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.

[26]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.

2. In virtue of the practical stakes, she doesn't believe that *p*;
3. In virtue of the practical stakes, she doesn't justifiably believe that *p*, although she does actually believe it.
4. In virtue of the practical stakes, she doesn't know that *p*, although she does justifiably believe it.

I think option 1 is implausibly sceptical, at least if applied to all cases like Coraline's. I've said that the probability of *p* is 0.99, but it should be clear that all that matters to generating a case like this is that *p* is not completely certain. Unless knowledge requires certainty, we'll be able to generate Coraline-like cases where there is sufficient evidence for knowledge. So that's ruled out.

Option 2 is basically what the Doxastic IRI theorist has to say. If Coraline has enough evidence to know *p*, but doesn't know *p* due to practical stakes, then the Doxastic IRI theorist is committed to saying that the practical stakes block *belief* in *p*. That's the Doxastic IRI position; stakes matter to knowledge because they matter to belief.

But that's also an implausible description of Coraline's situation. She is very confident that *p*. Her confidence is grounded in the evidence in the right way. She is insensitive in her actual deliberations to the difference between her evidence for *p* and evidence that guarantees *p*. She would become sensitive to that difference if someone offered her a bet that she knew was a 1000-to-1 bet on *p*, but she doesn't know that's what is on offer. In short, there is no difference between her unconditional attitudes, and her attitudes conditional on *p*, when it comes to any live question. That's enough, I think, for belief. So she believes that *p*. And that's bad news for the Doxastic IRI theorist; since it means here that stakes matter to knowledge without mattering to belief. I conclude, reluctantly, that Doxastic IRI is false.

3.3 Stakes as Defeaters

That still leaves two options remaining, what I've called options 3 and 4 above. Option 3, if suitably generalised, says that knowledge is practically sensitive because the justification condition on belief is practically sensitive. Option 4 says that practical considerations impact knowledge directly. As I read them, Jeremy Fantl and Matthew McGrath defend a version of Option 3. In the next and last subsection, I'll argue against that position. But first I want to sketch what a position like option 4 would look like.

Knowledge, unlike justification, requires a certain amount of internal coherence among mental states. Consider the following story from David Lewis:

> I speak from experience as the repository of a mildly inconsistent corpus. I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (Lewis, 1982, 436)

I think in that case that Lewis doesn't know that Nassau Street runs roughly east-west. (From here on, call the proposition that Nassau Street runs roughly east-west *N*.) If his belief that it does was acquired and sustained in a suitably reliable way,

then he may well have a justified belief that N. But the lack of coherence with the rest of his cognitive system, I think, defeats any claim to knowledge he has.

Coherence isn't just a requirement on belief; other states can cohere or be incoherent. Assume Lewis corrects the incoherence in his beliefs, and drops the belief that Nassau Street the railway are roughly parallel. Still, if Lewis believed that N, preferred doing φ to doing ψ conditional on N, but actually preferred doing ψ to doing φ, his cognitive system would also be in tension. That tension could, I think, be sufficient to defeat a claim to know that N.

And it isn't just a requirement on actual states; it can be a requirement on rational states. Assume Lewis believed that N, preferred doing φ to doing ψ conditional on N, and preferred doing φ to doing ψ, but should have preferred doing ψ to doing φ given his interests. Then I think the fact that the last preference is irrational, plus the fact that were it corrected there would be incoherence in his cognitive states defeats the claim to know that N.

A concrete example of this helps make clear why such a view is attractive, and why it faces difficulties. Assume there is a bet that wins \$2 if N, and loses \$10 if not. Let φ be taking that bet, and ψ be declining it. Assume Lewis shouldn't take that bet; he doesnt have enough evidence to do so. Then he clearly doesn't know that N. If he knew that N, φ would dominate ψ, and hence be rational. But it isn't, so N isn't known. And that's true whether Lewis's preferences between φ and ψ are rational or irrational.

Attentive readers will see where this is going. Change the bet so it wins a penny if N, and loses \$1,000 if not. Unless Lewis's evidence that N is incredibly strong, he shouldn't take the bet. So, by the same reasoning, he doesn't know that N. And we're back saying that knowledge requires incredibly strong evidence. The solution, I say, is to put a pragmatic restriction on the kinds of incoherence that matter to knowledge. Incoherence with respect to irrelevant questions, such as whether to bet on N at extremely long odds, doesn't matter for knowledge. Incoherence (or coherence obtained only through irrationality) does. The reason, I think, that Non-Doxastic IRI is true is that this coherence-based defeater is sensitive to practical interests.

The string of cases about Lewis and N has ended up close to the Coraline example. We already concluded that Coraline didn't know p. Now we have a story about why - belief that p doesn't cohere sufficiently well with what she should believe, namely that it would be wrong to take the bet. If all that is correct, just one question remains: does this coherence-based defeater also defeat Coraline's claim to have a justified belief that p? I say it does not, for three reasons.

First, 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 justified.

Second, 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 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 justified, but when we change tools, and start using beliefs, the state is unjustified. Given this outlook on beliefs and credences, it is natural to say that her belief is justified. Natural, but not compulsory, for reasons Jeremy Fantl pointed out to me.[27] We don't want a metaphysics on which persons and philosophers are separate entities. Yet we can say that someone is a good person but a bad philosopher. Normative statuses can differ depending on which property of a thing we are considering. That suggests it is at least coherent to say that one and the same state is a good credence but a bad belief. But while this may be coherent, I don't think it is well motivated, and it is natural to have the evaluations go together.

Third, we don't *need* to say that Coraline's belief in p is unjustified in order to preserve other nice theories, in the way that we do need to say that she doesn't know p in order to preserve a nice account of how we understand decision tables. It's this last point that I think Fantl and McGrath, who say that the belief is unjustified, would reject. So let's conclude with a look at their arguments.

3.4 Fantl and McGrath on Interest-Relativity

Fantl and McGrath's argue for the principle (JJ), which entails that Coraline is not justified in believing p.

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

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

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

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. But I think it is false, and the Coraline example shows it is false. Coraline isn't justified in taking the bet, and is justified in believing p, but more evidence for p would suffice for taking the bet. So Coraline's case shows that (JJ) is false. But there are a number of possible objections to that position. I'll spend the rest of this section, and this paper, going over them.[28]

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

---

[27]The following isn't Fantl's example, but I think it makes much the same point as the examples he suggested.

[28]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/.

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 than 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 with the best outcome is an *ex ante* unjustifiable risk. Call that action φ, and let $B\varphi$ be the proposition that φ has the best outcome. Then $B\varphi$ is true, and conditional on $B\varphi$, φ has an excellent expected return. But doing φ 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.

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.

Having said 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 hence justifiably believes taking the bet will return the best outcome, and knows all this, but still 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 (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: 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*.[29]

Reply: This objection this one turns on an equivocation 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. 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.

Once we have an expansive theory of defeaters, as I've adopted here, it becomes problematic to describe the case in the language Fantl and McGrath use. They focus a lot on whether agents like Coraline have 'knowledge-level justification' for *p*, which is defined as "justification strong enough so that shortcomings in your strength of justification stand in the way of your knowing". (Fantl and McGrath, 2009, 97). An important part of their argument is that an agent is justified in believing *p* iff they have knowledge-level justification for *p*. I haven't addressed this argument, so I'm not really addressing the case on their terms.

---

[29]Compare the 'subtraction argument' on page 99 of Fantl and McGrath (2009).

Well, does Coraline have knowledge-level justification for *p*? I'm not sure, because I'm not sure I grasp this concept. Compare the agent in Harman's dead dictator case (Harman, 1973, 75). Does she have knowledge-level justification that the dictator is dead? In one sense yes; it is the existence of misleading news sources that stops her knowing. In another sense no; she doesn't know, but if she had better evidence (e.g., seeing the death happen) she would know. I want to say the same thing about Coraline, and that makes it hard to translate the Coraline case into Fantl and McGrath's terminology.

References

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. (11)

Braddon-Mitchell, David and Jackson, Frank, (2007). *The Philosophy of Mind and Cognition, second edition*. Malden, MA: Blackwell. (13)

Brown, Jessica, (2008). "Knowledge and Practical Reason." *Philosophy Compass* 3: 1135-1152, doi:10.1111/j.1747-9991.2008.00176.x. (19)

Cohen, Stewart, (1988). "How to be a Fallibilist." *Philosophical Perspectives* 2: 91-123, doi:10.2307/2214070. (1)

DeRose, Keith, (1995). "Solving the Skeptical Problem." *Philosophical Review* 104: 1-52, doi:10.2307/2186011. (1)

Fantl, Jeremy and McGrath, Matthew, (2009). *Knowledge in an Uncertain World*. Oxford: Oxford University Press. (10, 24, 27)

Feltz, Adam and Zarpentine, Chris, (2010). "Do You Know More When It Matters Less?" *Philosophical Psychology* 23: 683-706, doi:10.1080/09515089.2010.514572. (11)

Gettier, Edmund L., (1963). "Is Justified True Belief Knowledge?" *Analysis* 23: 121-123, doi:10.2307/3326922. (27)

Hammond, Peter J., (1988). "Consequentialist Foundations for Expected Utility." *Theory and Decision* 25: 25-78, doi:10.1007/BF00129168. (15)

Harman, Gilbert, (1973). *Thought*. Princeton: Princeton University Press. (28)

Hawthorne, John, (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press. (8)

Hawthorne, John and Stanley, Jason, (2008). "Knowledge and Action." *Journal of Philosophy* 105: 571-590, doi:10.5840/jphil20081051022. (10)

Ichikawa, Jonathan, (2009). "Explaining Away Intuitions." *Studia Philosophica Estonica* 22: 94-116, doi:10.12697/spe.2009.2.2.06. (11)

Jackson, Frank, (1991). "Decision Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461-482, doi:10.1086/293312. (17)

Kennett, Jeanette and Smith, Michael, (1996). "Frog and Toad Lose Control." *Analysis* 56: 63-73, doi:10.1111/j.0003-2638.1996.00063.x. (26)

—, (1996). "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. (26)

Lewis, David, (1969). *Convention: A Philosophical Study*. Cambridge: Harvard University Press. (7)

—, (1982). "Logic for Equivocators." *Noûs* 16: 431-441, doi:10.1017/cbo9780511625237.009. Reprinted in *Papers in Philosophical Logic*, pp. 97-110. (22)

—, (1996). "Elusive Knowledge." *Australasian Journal of Philosophy* 74: 549-567, doi:10.1080/00048409612347521. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 418-446. (1)

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. (26)

Runyon, Damon, (1992). *Guys & Dolls: The stories of Damon Runyon*. New York: Penguin. (9)

Stalnaker, Robert, (2008). *Our Knowledge of the Internal World*. Oxford: Oxford University Press. (1)

Stanley, Jason, (2005). *Knowledge and Practical Interests*. Oxford University Press. (3)

Sturgeon, Scott, (2008). "Reason and the Grain of Belief." *Noûs* 42: 139-165, doi:10.1111/j.1468-0068.2007.00676.x. (14)

Watson, Gary, (1977). "Skepticism about Weakness of Will." *Philosophical Review* 86: 316-339, doi:10.2307/2183785. (26)

Weatherson, Brian, (2003). "What Good Are Counterexamples?" *Philosophical Studies* 115: 1-31, doi:10.1023/A:1024961917413. (6)

—, (2005). "Can We Do Without Pragmatic Encroachment?" *Philosophical Perspectives* 19: 417-443, doi:10.1111/j.1520-8583.2005.00068.x. (2, 3, 12, 13, 14, 16, 26)

—, (2006). "Questioning Contextualism." In Stephen Cade Hetherington (ed.), *Epistemology Futures*, 133-147. Oxford: Oxford University Press. (11)

Williamson, Timothy, (1998). "Conditionalizing on Knowledge." *British Journal for the Philosophy of Science* 49: 89-121, doi:10.1093/bjps/49.1.89. (26)

—, (2000). *Knowledge and its Limits*. Oxford University Press. (6, 7, 26)