Knowledge, Bets and Interests

Brian Weatherson

---

1 Knowledge in Decision Making

In my (2005), I defended three claims. Roughly, they were:

(1) Whether a person believes that p depends on their interests, broadly construed.
(2) Whether a person knows that p depends on their interests, broadly construed.
(3) The explanation for why (2) is true is simply that (1) is true; we don't need to add any interest-relativity to epistemology that isn't already present in our theory of belief.

The focus in that paper was on justified belief, and in particular that the only reason justified belief is interest-relative is that belief is interest-relative. But I did strongly suggest that I thought the same was true of knowledge. In this paper I want to give a new argument for (2), and argue against (3). That is, I now think that some of the interest-relativity of knowledge cannot be explained by the interest-relativity of belief. This is largely because I'm now convinced by a version of the example that Jason Stanley (2005) calls 'Ignorant High Stakes'. It will be easier to see why this case is a problem for the package of views I earlier adopted by looking at the new argument for (2). In the earlier paper I argued that the best way to understand the relationship between belief and degree of belief was to see that belief was interest-relative.[1] I'll start this paper by arguing that the best way to understand our orthodox practices in decision theory leads to an interest-relative theory 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

---

[1] Scott Sturgeon (2008) offers an excellent treatment of how puzzling this relationship is. But I think Sturgeon doesn't pay sufficient attention to the possibilities for interest-relative treatments of the relationship, and so is led to adopt a sub-optimal resolution of the puzzles he raises.

for the students to think that the states are dependent on the choices in either way, and in fact the students don't worry about that kind of dependence.

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

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

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

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

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

I've left Alice's story fairly schematic; let's fill in some of the details. Alice is on holiday at a casino. It's a fair casino; the probabilities of the outcomes of each of the games is just what you'd expect. And Alice knows this. The table she's standing at is a roulette table. The token is a chip from the casino worth $1000. Putting the token on the table means placing a bet. As it turns out, it means placing a bet on the roulette wheel landing on 28. If that bet wins she gets her token back and another token of the same value. There are many other bets she could make, but Alice has decided not to make all but one of them. Since her birthday is the 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.[2] So the above table, which makes it look like placing the bet is the dominant, and hence rational, option, is misleading.

---

[2]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.

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

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

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

I'm here retracting some things I said a few years ago in a paper on philosophical methodology (Weatherson, 2003). There I argued that identifying knowledge with justified true belief would give us a theory on which knowledge was more natural than a theory on which we didn't identify knowledge with any other epistemic property. I now think that is wrong for a couple of reasons. First, although it's true (as I say in the earlier paper) that knowledge can't be primitive or perfectly natural, this 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.[3]

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

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

The argument here that the 'special doxastic relationship' is knowledge is clearly far from conclusive. But it is plausible, I think, that the relationship should be a fairly simple, natural relationship. And it seems that any simple, natural relationship weaker than knowledge will be so weak that when we plug it into our decision theory, it will say that Alice should do clearly irrational things in one or other of the cases we described above. And it seems that any simple, natural relationship stronger than knowledge will be so strong that it makes decision theory or game theory impractical.

---

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

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

1.2   From Decision Theory to Interest-Relativity

This way of thinking about decision problems offers a new perspective on the issue of whether we should always be prepared to bet on what we know.[4] To focus intuitions, let's take a concrete case. Barry is sitting in his apartment one evening when he hears a musician performing in the park outside. The musician, call her Beth, is one of Barry's favourite musicians, so the music is familiar to Barry. Barry is excited that Beth is performing in his neighbourhood, and he decides to hurry out to see the show. As he prepares to leave, a genie appears an offers him a bet. If he takes the bet, and the musician is Beth, then the genie 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. Including it is like including a column for what will happen if the genie is lying about the consequences of taking or declining the bet. So let's write the table in the standard form.

	Musician is Beth
Take Bet	Win $10
Decline Bet	Status quo

---

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

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

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

1. Barry never knew that the musician was Beth.

2. Barry did know that the musician was Beth, but this knowledge was destroyed by the genie's offer of the bet.

3. States of the world that are known not to obtain should still be represented in decision problems, so taking the bet is not a dominating option.

The first option is basically a form of scepticism. If the take-away message from the above discussion is that Barry doesn't know the musician is Beth, we can mount a similar argument to show that he knows next to nothing.[5] And the third option would send us back into the problems about interpreting and applying decision theory that we spend the first few pages trying to get out of.

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.

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. As we'll see later in the paper, I think 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.

The case of Barry and Beth 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

---

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

of interest-relativity about knowledge. That is, my view is a form of interest-relative-invariantism, or IRI.[6]

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 argument for interest-relativity in (Weatherson, 2005) relied heavily on the relationship between degrees of belief, or credences, and full belief. In particular, it relied on looking at the functional roles of both credences and full belief. Frank Ramsey (1926) provides a clear statement of one of the key functional roles of credences; their connection to action. Of course, Ramsey did not take himself to be providing one component of the functional theory of credence. He took himself to be providing a behaviourist/operationalist reduction of credences to dispositions. But we do not have to share Ramsey's metaphysics to use his key ideas. Those ideas include that it's distinctively *betting* dispositions that are crucial to the account of credence, and that all sorts of actions in everyday life constitute bets.

The connection to betting behaviour lives on today most prominently in the work on 'representation theorems'.[7] What a representation theorem shows is that for any agent whose pairwise preferences satisfy some structural constraints, there is a probability function and a utility function such that the agent prefers bet X to bet Y just in case the expected utility of X (given that probability and utility function) is greater than that of Y. Moreover, the probability function is unique (and the utility function is unique up to positive affine transformations). Given that, it might seem plausible to identify the agent's credence with this probability function, and the agent's (relative) values with this utility function.

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

---

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

[7]See Maher (1993) for the most developed account in recent times.

is in a particular credal state, we have to consider not just their actions, and their disposition to act, but the connections between the alleged credal state and other states.

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

When writers start to think about the connection between belief and credence, they run into the following problem fairly quickly. The alleged possibility where S believes p, and doesn't believe q, but her credence in q is higher than her credence in p strikes many theorists as excessively odd.[8] That suggests the so-called 'threshold view' of belief, that belief is simply credence above a threshold. It is also odd to say that a rational, reflective agent could believe p, believe q, yet take it as an open question whether $p \wedge q$ is true, refusing to believe or disbelieve it. Finally, it seems we can believe things to which we don't give credence 1. In the case of Barry and Beth from section 1, for example, before the genie comes in, it seems Barry does believe the musician is Beth. But he doesn't have credence 1 in this, since having credence 1 means being disposed to make a bet at any odds.

We can raise the same kind of problem by looking directly at functional roles. A key functional role of credences is that if an agent has credence x in p she should be prepared to buy a bet that returns 1 util if p, and 0 utils otherwise, iff the price is no greater than x utils. A key functional role of belief is that if an agent believes p, and recognises that 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 as we argued in the previous paragraph, it is wrong to identify belief with credence 1.

If we spell out more carefully what the functional states of credence and belief are, a loophole emerges in the argument that belief implies credence 1. The interest-relative theory of belief exploits that loophole. What's the difference, in functional terms, between having credence x in p, and having credence $x + \varepsilon$ in p? Well, think again about the bet that pays 1 util if p, and 0 utils otherwise. And imagine that bet is offered for $x + \varepsilon/2$ utils. The person whose credence is x will decline the offer; the person whose credence is $x + \varepsilon$ will accept it. Now it will usually be that no such bet is on offer.[9] No matter; as long as one agent is *disposed* to accept the offer, and the

---

[8] I'm going to argue below that cases like that of Barry and Beth suggest that in practice this isn't nearly as odd as it first seems.

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

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

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

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

This suggests the key difference between belief and credence 1. A credence in p of 1, in a rational agent, 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.[10]

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 $U(\phi|p) \geq U(\psi|p)$.

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.

If we make the standard kinds of assumptions in orthodox decision theory, i.e., assume at least some form of probabilism and consequentialism[11], then the agent will answer each of these questions the same way simpliciter and conditional on p.

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

---

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

[11]I mean here consequentialism in roughly the sense used by Hammond (1988).

- Is it better to do ϕ or ψ?

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

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

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

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

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

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

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

other questions become relevant in the course of her inquiry into this question.[12] Charlie decides that to the first decimal place, $\Pr(p) = 1.0$, i.e., $\Pr(p) > 0.95$. That is compatible with simply believing that p. And that seems right; if for practical purposes, the probability of p is indistinguishable from 1, then the agent is confident enough in p to believe it.

2.2 A Worked Example

To get a feel for how this theory works in practice, it's helpful to go through a particular case, such as one that is alleged by Ram Neta to be hard for interest-relative theorists to accommodate.

> Kate needs to get to Main Street by noon: her life depends upon it. She is desperately searching for Main Street when she comes to an intersection and looks up at the perpendicular street signs at that intersection. One street sign says "State Street" and the perpendicular street sign says "Main Street." Now, it is a matter of complete indifference to Kate whether she is on State Street–nothing whatsoever depends upon it. (Neta, 2007, 182)

Let's assume for now that Kate is rational; dropping this assumption introduces mostly irrelevant complications.[13] Kate will not believe she's on Main Street. She would only have that belief if she took it to be settled that she's on Main, and hence not worthy of spending further effort investigating. But presumably she won't do that. The rational thing for her to do is to get confirming (or if relevant confounding) evidence for the appearance that she's on Main. If it were settled that she was on Main, the rational thing to do would be to try to relax, and be grateful that she had found Main Street. Since she has different attitudes about what to do *simpliciter* and conditional on being on Main Street, she doesn't believe she's on Main Street.

So far so good, but what about her attitude towards the proposition that she's on State Street? She has enough evidence for that proposition that her credence in it should be rather high. And no practical issues turn on whether she is on State. So she believes she is on State, right?

Not so fast! Believing that she's on State has more connections to her cognitive system than just producing actions. It's true that Kate should, and will, act exactly as if she were on State Street. But that's not enough for belief. It is only rational to believe she is on State Street if the sign she's looking at is accurate. It's possible that the sign is only partially accurate, and she is on State but not Main. But unless the sign is accurate, any belief that she's on State is undersupported. And Kate knows this. So belief that she's on State stands and falls with belief that the sign is accurate. But forming a belief that the sign is accurate, i.e., settling the question of whether the sign is accurate in the affirmative, means she should form the belief that she's

---

[12]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.

[13]Neta's own treatment of the case implies, by my lights, that Kate is irrational, since he thinks that Kate does believe that she's on Main Street on this basis. I say more about how the theory applies to irrational agents in section 4.1, and what I say there would apply equally well to Kate's case if she is irrational.

on Main. And that's not something she should believe, for reasons we've gone into above.

So she shouldn't believe that she's on State, because that belief is tied up too closely to belief in a proposition that she shouldn't take to be settled. If she does believe she's on State, she would have an irrational attitude towards that proposision, and that kind of irrationality would be inconsistent with knowledge. So she doesn't know, and can't know, that she's either on State or on Main.

Neta thinks that the best way for the interest-relative theorist to handle this case is to say that the high stakes associated with the proposition that Kate is on Main Street imply that certain methods of belief formation do not produce knowledge. And he argues, plausibly, that such a restriction will lead to implausibly sceptical results. But that's not the right way for the interest-relative theorist to go. What they should say, and what I do say, is that Kate can't know she's on State Street because the only grounds for that belief is intimately connected to a proposition that, in virtue of her interests, she needs very large amounts of evidence to believe.

The take-home lesson from this is that inferential connections between propositions matter a lot on the functionalist/interest-relative theory that I'm presenting. These inferential connections can come and go quickly. Imagine that Kate sees a glimpse of a food truck that looks a lot like the cupcake truck. She thinks to herself, "It would be great to go there and get a cupcake, but I have to run, I need to be on Main Street by noon." She then glances at her phone and sees a message that the cupcake truck is on Main Street. Once she gets that information the question of whether the truck she's looking at is the cupcake truck is inferentially tied to the question of whether the street she's on is Main Street. She can't form the belief that the truck is the cupcake truck without settling the question of whether she's on Main Street. And a glimpse of a food truck is not enough information to settle a life-or-death question, like the question she is facing. So she should cease believing that the truck is indeed the cupcake truck, even though the epistemic probability that the truck is the cupcake truck has not fallen, and there's no reason her credence that the truck is the cupcake truck has fallen either.

This all seems like the right thing to say about the case to me. Admittedly the case is difficult, and intuitions about it aren't of great theoretical weight. But to the extent that the intuitions are of any weight, they tell against the view that beliefs should be identified with credence above a threshold. In that example, whether Kate believes a proposition can change without her credence in the proposition changing. That's inconsistent with belief just being credence above a fixed threshold.

2.3 Two Caveats

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

Jones has a false geographic belief: he believes that Los Angeles is west of Reno, Nevada.[14] 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. First, we say that which questions are relevant questions is not just relative to the agent's interests, but also relevant to the proposition being considered. A question may be relevant relative to p, but not relative to q. Second, we say that relative to p, the question of whether to believe p is a relevant question. Third, we say that an agent only prefers believing p to not believing it if their credence in p is greater than their credence in $\neg p$, i.e., if their credence in p is greater than $1/2$. Finally, we say that when the issue is whether the subject believes that p, the question of whether to believe p is not just a relevant question on its own, but it stays being a relevant question conditional on any q that is relevant to the subject. In the earlier paper (Weatherson, 2005) I argue that this solves the problem raised by impractical propositions in a smooth and principled way.

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

---

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

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, $1/2$. 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 $Pr(p) = 1/2$, 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.[15]

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.

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

---

[15]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.

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.[16]

> 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 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

---

[16]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.

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   Can Interest-Relativity Be Isolated?

3.1   Fantl and McGrath on Interest-Relativity

Jeremy Fantl and Matthew McGrath (2009) have argued that my interest-relative theory of belief cannot explain all of the interest-relativity in epistemology. I'm going to agree with their conclusion, but not with their premises. And this isn't just a dispute about arguments. I think that the interest-relative theory of belief *can* explain (just about) all the interest-relativity in the theory of justified belief.[17]

Fantl and McGrath's primary complaint against the interest-relative theory of belief is that it is not strong enough to entail principles such as (JJ).

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

---

[17]In the earlier paper I said that a justified belief was a justified credence that is high enough to constitute a belief. The only normative concept in that definition is that of a justified credence, which I argued was not interest-relative. It's arguable that if a credence is not perfectly justified, then *how* unjustified the resulting belief is will be interest-relative, which is why I said 'just about' in the main text. But exploring this would take us too far away from the discussion of knowledge which is the focus of this volume.

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

- She knows that p and q are independent, so her credence in any conjunction where one conjunct is a member of $\{p, \neg p\}$ and the other is a member of $\{q, \neg q\}$ will be the product of her credences in the conjuncts.
- Her credence in p is 0.99, just as the evidence supports.
- Her credence in q is also 0.99. This is unfortunate, since the rational credence in q given her evidence is 0.01.
- She has a choice between taking and declining a bet with the following payoff structure.[18] (Assume that the marginal utility of money is close enough to constant that expected dollar returns correlate more or less precisely with expected utility returns.)

	$p \wedge q$	$p \wedge \neg q$	$\neg p$
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.[19] That is, had the expected utility given her credences been close to 0, she would have not acted until she made a computation. But from her perspective this looks like basically a free $100, so she takes it. Happily, this all turns out well enough, since p is true. But it was a dumb thing to do. The expected utility of taking the bet given her evidence is negative, it is a little under -$8. So she isn't warranted, given her evidence, in taking the bet.

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

1. p is not justified enough to warrant her in taking the bet.
2. She believes p.[20]

---

[18]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.

[19]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.

[20]In terms of the example discussed in the previous footnote, she believes that the bank will survive, i.e., that she'll get her money back if she deposits it.

3. This belief is rational.

The argument for 1 is straightforward. She isn't warranted in taking the bet, so p isn't sufficiently warranted to justify it. This is despite the fact that p is obviously relevant. Indeed, given p, taking the bet strictly dominates declining it. But still, p doesn't warrant taking this bet, because nothing warrants taking a bet with negative expected utility. Had the rational credence in p been higher, then the bet would have been reasonable. Had the reasonable credence in p been, say, 0.9999, then she would have been reasonable in taking the bet, and using p as a reason to do so. So there's a good sense in which p simply isn't warranted enough to justify taking the bet.[21]

The argument for 2 is that she has a very high credence in p, this credence is grounded in the evidence in the right way, and it leads her to act as if p is true, e.g. by taking the bet. It's true that her credence in p is not 1, and if you think credence 1 is needed for belief, then you won't like this example. But if you think that, you won't think there's much connection between (JJ) and pragmatic conditions in epistemology either. So that's hardly a position a defender of Fantl and McGrath's position can hold.[22]

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

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

So that's the argument that (JJ) is false. And if it's false, the fact that the interest-relative theory doesn't entail it is a feature, not a bug. But there are a number of possible objections to that position. I'll spend the rest of this section going over them.[23]

---

[21]I think this is exactly the sense in which 'is warranted enough' is being used in (JJ), though I'm not entirely sure about this. For present purposes, I plan to simply interpret (JJ) that way, and not return to exegetical issues.

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

[23]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/.

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

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

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

It might be controversial to use this very case to reject premise 1. But the falsity of premise 1 should be clear on independent grounds. What p entails is that Coraline will be best off by taking the bet. But there are lots of things that will make me better off that I shouldn't do. Imagine I'm standing by a roulette wheel, and the thing that will make me best off is betting heavily on the number 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 $B\phi$ be the proposition that ϕ has the best outcome. Then $B\phi$ is true, and conditional on $B\phi$, 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.

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

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

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

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

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

The objector might think I am being uncharitable here. The objection doesn't say that Coraline's knowledge provides an *inductive* reason to take the bet. Rather, they say, it provides a *conclusive* reason to take the bet. And conclusive reasons cannot be defeated by irrational beliefs elsewhere in the web. Here we reach an impasse. I say that knowledge that you justifiably believe p cannot provide a conclusive reason to bet on p because I think Coraline knows she justifiably believes p, but does not have a conclusive reason to bet on p. That if, I think the premise the objector uses here is false because I think (JJ) is false. The person who believes in (JJ) won't be so impressed by this move.

Having said all that, the more complicated example at the end of Weatherson (2005) was designed to raise the same problem without the consequence that if p is

true, the bet is sure to return a positive amount. In that example, conditionalising on p means the bet has a positive expected return, but still possibly a negative return. But in that case (JJ) still failed. If accepting there are cases where an agent justifiably believes p, and knows this, but can't rationally bet on p is too much to accept, that more complicated example might be more persuasive. Otherwise, I concede that someone who believes (JJ) and thinks rational agents can use it in their reasoning will not think that a particular case is a counterexample to (JJ).

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

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

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

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

Stepping back a bit, there's a reason the interest-relative theory says that the ideal and justification come apart right here. On the interest-relative theory, like on any pragmatic theory of mental states, the *identification* of mental states is a somewhat holistic matter. Something is a belief in virtue of its position in a much broader network. But the *evaluation* of belief is (relatively) atomistic. That's why Coraline is justified in believing p, although if she were wiser she would not believe it. If she

were wiser, i.e., if she had the right attitude towards q, the very same credence in p would not count as a belief. Whether her state counts as a belief, that is, depends on wide-ranging features of her cognitive system. But whether the state is justified depends on more local factors, and in local respects she is doing everything right.

Objection: Since the ideal agent in Coraline's position would not believe p, it follows that there is no *propositional* justification for p. Moreover, doxastic justification requires propositional justification[24] So Coraline is not doxastically justified in believing p. That is, she isn't justified in believing p.

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

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

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

Objection: If Coraline is justified in believing p, then Coraline can use p as a premise in practical reasoning. If Coraline can use p as a premise in practical reasoning, and p is true, and her belief in p is not Gettiered, then she knows p. By hypothesis, her belief is true, and her belief is not Gettiered. So she should know p. But in the previous section it was argued that she doesn't know p. So by several steps of modus tollens, she isn't justified in believing p.[25]

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

---

[24]See Turri (2010) for a discussion of recent views on the relationship between propositional and doxastic justification. This requirement seems to be presupposed throughout that literature.

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

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.

3.2   Pragmatic Defeaters

As I mentioned at the top, Jason Stanley thinks that cases like his **Ignorant High Stakes** are important arguments for the importance of interest-relativity to the theory of knowledge. Moreover, he thinks that this shows that knowledge is interest-relative in a way that doesn't merely reflect the interest relativity of belief. I now think that cases like Coraline show that is correct. It is, I guess, a little misleading to describe Coraline as *ignorant* of the stakes. But she does have a false belief about the stakes, or at least about the odds.*I'll return towards the end to the question of whether stakes or odds are really crucial. It should be somewhat obvious that on my theory it is odds, not stakes, that matter, and I'll say a bit more about the benefits to framing the theory that way.* Hence she doesn't know what the odds are. Hence she is ignorant of the odds, at least in the sense of lacking knowledge.

The reason Coraline's case is a problem is that if we say she knows p, and we have established that knowledge structures decision problems, then the decision table Coraline faces looks like this.

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

Now taking the bet dominates declining the bet. So Coraline should take the bet, if this table is correct. But she shouldn't take the bet. So the table isn't correct. So Coraline must not know p.

But what could be the explanation of Coraline's lack of knowledge? It isn't that she doesn't believe p. We just showed that she does. It isn't even that she lacks a *justified* belief in p. We just showed that she has that too. It must be that her mistaken views about q somehow *defeat* her claim to knowledge that p. But it's not generally true that having odd views about q defeats knowledge that p. It's only because of her interests, i.e., because she faces a bet that is sensitive to both p and q, that mistaken views about q defeat knowledge that p. Hence her lack of knowledge here is dependent, in a crucial way, on her interests.

In slogan form, the conclusion of the previous paragraph is *There are pragmatic defeaters*. The cases where these arise are admittedly rare. But sometimes an agent who has a justified true belief that p doesn't know that p, because of the way p is tied up with other choices that she is currently facing.

Conceptualising the influence of interests on knowledge as coming via defeaters helps see what's mistaken about some recent criticisms of interest-relative invariantism, or IRI.[26] To see the potential problem here, imagine that there is another person, call her Dora, who has the same evidence and credences as Coraline, but who does not have the option of taking this bet. According to my version of IRI, she knows that p. Michael Blome-Tillmann (2009a) suggests this is implausible. As he notes, it means we can say things like (1).[27]

(1) Coraline and Dora have exactly the same evidence for p, but only Dora has enough evidence to know p, Coraline doesn't.

Blome-Tillman claims this is intuitively false, and hence IRI, which entails it, is false.

The first thing we might note here is that there's an interesting ambiguity in 'enough' that makes it tricky to say whether IRI really entails that (1) is true. Compare the following situation.

> George and Ringo both have $6000 in their bank accounts. They both are thinking about buying a new computer, which would cost $2000. Both of them also have rent due tomorrow, and they won't get any more money before then. George lives in New York, so his rent is $5000. Ringo lives in Syracuse, so his rent is $1000.

In that story, is (2) true or false?

(2) George and Ringo have the same amount of money in their bank accounts, but only Ringo has enough money to buy the computer, George does not.

I think the right thing to say is probably that (2) is ambiguous, with the ambiguity turning on just how we interpet 'enough'. In one sense, George does have enough money to buy the computer: he has $6000 in his bank account, and the computer costs $2000. So in that sense, (2) is false. In another (perhaps less preferred) sense, George does not have enough money to buy the computer. It is essential that he pays his rent, and if he buys the computer he won't be able to pay the rent. In that sense, (2) is true. But in that sense, it isn't surprising that (2) is true, since what one has enough money to buy, in that sense, depends not just on the size of one's bank account, but on one's forthcoming debts. These debts defeat George's claim to have enough money to buy the computer, even though he has the same amount of money as Ringo in his bank account.

We can see the same kind of situation arise in distinctively epistemic puzzle cases. The following case is a variant on the 'dead dictator' case developed by Gilbert Harman (1973).

---

[26]Of course, I don't think all, or even most, of the influence of interests on knowledge comes via defeaters. I think it mostly comes via the interest-relativity of belief. But the focus here is on the impact interests make on knowledge that isn't mediated by the impact they make on belief.

[27]Compare his example about John and Paul on page 329.

George and Ringo are scientists investigating whether more than 50% of Fs are Gs. They plan to run a barrage of tests on this. The first is just to collect a large random sample of Fs, and see whether they are Gs. To save money, they get a third party to do the data collection. The report says that over 60% of the sampled Fs are Gs, and the sample is large enough that this alone is enough to ground knowledge that more than 50% of Fs are Gs, and both of them rationally form the belief that more than 50% of Fs are Gs. But they both have large grants, and plan to run the other tests. It will turn out that Ringo's tests confirm the fact that more than 50% of Fs are Gs. But George's tests will, very misleadingly, indicate that the percentage of Fs that are Gs is actually somewhat under 50, and that the sample they originally used was not genuinely random.

After they've got the sample, but before they run the later tests, I think that Ringo knows that more than 50% of Fs are Gs, but George does not. As in Harman's case, the misleading evidence surrounding George defeats his putative knowledge. But Ringo is not, in the relevant sense, surrounded by that evidence; he was never going to collect it until after he had so much evidence that he'd know to dismiss it as misleading.

Given that, is (3) true or false?

(3) George and Ringo have exactly the same evidence that more than 50% of Fs are Gs, but only Ringo has enough evidence to know that, George doesn't.

As should be expected by this stage, I think (3) is ambiguous. On the most natural disambiguation, it is false. George has enough evidence to know that more than 50% of Fs are Gs, just like Coraline has enough evidence to know p. It's just that both of their claims to knowledge are defeated by other circumstances.

There are readings though of both (1) and (3) on which they are both false. But the existence of these readings should not be counterintuitive once we have remembered (or reminded our intuitions) that defeaters exist. George needs more evidence to know because his existing evidence is defeated by the misleaders that surround him. Coraline needs more evidence to know because his existing evidence is defeated by her irrational attitude towards q, and attitude that matters because of her practical situation.

4 Challenges to Interest-Relative Invariantism

There have been a number of objections to IRI over the past few years. Some of these challenges I think can be met by any version of IRI. But some of them are best met by using the resources of the the theory offered here, one that starts with the interest-relativity of belief. We just saw one such challenge – Blome-Tillman's discussion of the supposed counterintuitiveness of denying that knowledge supervenes on evidence. We'll start with two other objections Blome-Tillman raises to IRI.

4.1 Temporal Embeddings

Michael Blome-Tillmann (2009a) has argued that tense-shifted knowledge ascriptions can be used to show that his version of Lewisian contextualism is preferable to Interest Relative Invariantism.[28]. His argument uses a variant of the well-known bank cases.[29] Let O be that the bank is open Saturday morning. If Hannah has a large debt, she is in a high-stakes situation with respect to O. She had in fact incurred a large debt, but on Friday morning the creditor waived this debt. Hannah had no way of anticipating this on Thursday. She has some evidence for O, but not enough for knowledge if she's in a high-stakes situation. Blome-Tillmann says that this means after Hannah discovers the debt waiver, she could say (4).

(4) I didn't know O on Thursday, but on Friday I did.

But I'm not sure why this case should be problematic for any version of Interest Relative Invariantism. As Blome-Tillmann notes, it isn't really a situation where Hannah's stakes change. She was never actually in a high stakes situation. At most her perception of her stakes change; she thought she was in a high-stakes situation, then realised that she wasn't. Blome-Tillmann argues that even this change in perceived stakes can be enough to make (4) true if Interest Relative Invariantism is true. Now actually I agree that this change in perception could be enough to make (4) true, but when we work through the reason that's so, we'll see that it isn't because of anything distinctive, let alone controversial, about Interest Relative Invariantism.

If Hannah is rational, then given her interests she won't be ignoring ¬O possibilities on Thursday. She'll be taking them into account in her plans. Someone who is anticipating ¬O possibilities, and making plans for them, doesn't know O. That's not a distinctive claim of Interest Relative Invariantism. Any theory should say that if a person is worrying about ¬O possibilities, and planning around them, they don't know O. And that's simply because knowledge requires a level of confidence that such a person simply does not show. If Hannah is rational, that will describe her on Thursday, but not on Friday. So (4) is true not because Hannah's practical situation changes between Thursday and Friday, but because her psychological state changes, and psychological states are relevant to knowledge.

If the version of Interest Relative Invariantism I've been defending is correct, then this is just what we should expect. It's possible for stakes to change what the subject knows without changing what the subject believes, but the cases where this happens are rare, typically involving irrational credences in somewhat related propositions. The standard kind of way in which the agent loses knowledge when the stakes rise is that she stops *believing* the target proposition.

---

[28]Blome-Tillmann calls Interest Relative Invariantism 'subject-sensitive invariantism'. This is an unfortunate moniker. The only subject-*insensitive* theory of knowledge has that for any S, T S knows that p iff T knows that p The view Blome-Tillmann is targetting certainly isn't defined in opposition to *this* generalisation.

[29]See Stanley (2005) for the versions that Blome-Tillman is building on. In the interests of space, I won't repeat them yet again here.

What if Hannah is, on Thursday, irrationally ignoring $\neg O$ possibilities, and not planning for them even though her rational self wishes she were planning for them? In that case, it seems she still believes O. After all, she makes the same decisions as she would as if O were sure to be true. It's true that she doesn't satisfy the canonical *input* conditions for believing O, but that's consistent with believing O. If functionalists didn't allow some deviation from optimal input conditions, there wouldn't be any irrational beliefs.

But it's worth remembering that if Hannah does irrationally ignore $\neg O$ possibilities, she is being irrational with respect to O. And it's very plausible that this irrationality defeats knowledge. That is, you can't be irrational with respect to a proposition and know it. Irrationality excludes knowledge. That may look a little like a platitude, so it's worth spending a little time on how it can lead to some quirky results for reasons independent of Interest Relative Invariantism.[30]

So consider Bobby. Bobby has the disposition to infer $\neg B$ from $A \rightarrow B$ and $\neg A$. He currently has good inductive evidence for q, and infers q on that basis. But he also knows $p \rightarrow q$ and $\neg p$. If he notices that he has these pieces of knowledge, he'll infer $\neg q$. This inferential disposition defeats any claim he might have to *know q*; the inferential disposition is a kind of doxastic defeater. Then Bobby sits down with some truth tables and talks himself out of the disposition to infer $\neg B$ from $A \rightarrow B$ and $\neg A$. He now knows q although he didn't know it earlier, when he had irrational attitudes towards a web of propositions including q. And that's true even though his evidence for q didn't change. I assume here that irrational inferential dispositions which the agent does not know he has, and which he does not apply, are not part of his evidence, but that shouldn't be controversial.

I think Bobby's case is just like Hannah's, at least under the assumption that Hannah simply ignores the significance of O to her practical deliberation. In both cases, defective mental states elsewhere in their cognitive architecture defeat knowledge claims. And in that kind of case, we should expect sentences like (4) to be true, even if they appear counterintuitive before we've worked through the details. The crucial point is that once we work through the details, we see that somewhat distant changes in the rest of the cognitive system changes what the agent knows. So, a little counterintuitively, it can be the case that an agent knows something after a distant change in the system, but not before. That's all that happens in Hannah's case.

4.2 Modal Embeddings

Blome-Tillman's other objection to IRI concerns modal embeddings of knowledge ascriptions. Jason Stanley had argued that the fact that IRI has counterintuitive consequences when it comes to knowledge ascriptions in modal contexts shouldn't count too heavily against IRI, because contextualist approaches are similarly counterintuitive. In particular, he argues that the theory that 'knows' is a contextually sensitive quantifier, plus the account of quantifier-domain restriction that he developed with

---

[30]There's a methodological point here worth stressing. Doing epistemology with imperfect agents often results in facing tough choices, where any way to describe a case feels a little counterintuitive. If we simply hew to intuitions, we risk being led astray by just focussing on the first way a puzzle case is described to us.

Zoltán Gendler Szabó (Stanley and Szabó, 2000), has false implications when it is applied to knowledge ascriptions in counterfactuals. Blome-Tillman disagrees, but I don't think he provides very good reasons for disagreeing. In fact, Stanley's argument seems to be a very good argument against Lewisian contextualism about 'knows'.[31] Let's start by reviewing how we got to this point.

Often when we say *All Fs are Gs*, we really mean *All C Fs are Gs*, where *C* is a contextually specified property. So when I say *Every student passed*, that utterance might express the proposition that every student **in my class** passed. Now there's a question about what happens when sentences like *All Fs are Gs* are embedded in various contexts. The question arises because quantifier embeddings tend to allow for certain kinds of ambiguity. For instance, when we have a sentence like *If p were true, all Fs would be G*, that could express either of the following two propositions. (We're ignoring context sensitivity for now, but we'll return to it in a second.)

- If *p* were true, then everything that would be *F* would also be *G*.
- If *p* were true, then everything that's actually *F* would be *G*.

We naturally interpret (5) the first way, and (6) the second way.

(5) If I had won the last Presidential election, everyone who voted for me would regret it by now.

(6) If Hilary Clinton had been the Democratic nominee in the last Presidential election, everyone who voted for Barack Obama would have voted for her.

Given this, you might expect that we could get a similar ambiguity with *C*. That is, when you have a quantifier that's tacitly restricted by *C*, you might expect that you could interpret a sentence like *If p were true, all Fs would be G* in either of these two ways. (In each of these interpretations, I've left *F* ambiguous; it might denote the actual *F*s or the things that would be *F* if *p* were true. So these are just partial disambiguations.)

- If *p* were true, then every *F* that would be *C* would also be *G*.
- If *p* were true, then every *F* that is actually *C* would be *G*.

Surprisingly, it's hard to get the second of these readings. Or, at least, it is hard to *avoid* the availability of the first reading. Typically, if we restrict our attention to the *C*s, then when we embed the quantifier in the consequent of a counterfactual, the restriction is to the things that would be *C*, not to the actual *C*s.[32]

Blome-Tillmann notes that Stanley makes these observations, and interprets him as moulding them into the following argument against Lewisian contextualism.

1. An utterance of *If p were true, all Fs would be Gs* is interpreted as meaning *If p were true, then every F that would be C would also be G*.

---

[31]The canonical source for Lewisian contextualism is Lewis (1996), and Blome-Tillmann defends a variant in Blome-Tillmann (2009b).

[32]See Stanley and Szabó (2000) and Stanley (2005) for arguments to this effect.

2. Lewisian contextualism needs an utterance of *If p were true, then S would know that q* to be interpreted as meaning *If p were true, then S's evidence would rule out all ¬q possibilities, except those that are actually being properly ignored*, i.e. it needs the contextually supplied restrictor to get its extension from the nature of the actual world.

3. So, Lewisian contextualism is false.

And Blome-Tillmann argues that the first premise of this argument is false. He thinks that he has examples which undermine premise 1. But I don't think his examples show any such thing. Here are the examples he gives. (I've altered the numbering for consistency with this paper.)

(7) If there were no philosophers, then the philosophers doing research in the field of applied ethics would be missed most painfully by the public.

(8) If there were no beer, everybody drinking beer on a regular basis would be much healthier.

(9) If I suddenly were the only person alive, I would miss the Frege scholars most.

These are all sentences of (more or less) the form *If p were true, Det Fs would be G*, where *Det* is some determiner or other, and they should all be interpreted a la our second disambiguation above. That is, they should be interpreted as quantifying over actual *F*s, not things that would be *F* if *p* were true. But the existence of such sentences is completely irrelevant to what's at issue in premise 1. The question isn't whether there is an ambiguity in the *F* position, it is whether there is an ambiguity in the *C* position. And nothing Blome-Tillmann raises suggests premise 1 is false. So this response doesn't work.

Even if a Lewisian contextualist were to undermine premise 1 of this argument, they wouldn't be out of the woods. That's because premise 1 is much stronger than is needed for the anti-contextualist argument Stanley actually runs. Note first that the Lewisian contextualist needs a reading of *If p were true, all Fs would be G* where it means:

- If *p* were true, every actual *C* that would be *F* would also be *G*.

The reason the Lewisian contextualist needs this reading is that on their story, *S knows that p* means *Every ¬p possibility is ruled out by S's evidence*, where the *every* has a contextual domain restriction, and the Lewisian focuses on the actual context. The effect in practice is that an utterance of *S knows that p* is true just in case every ¬p possibility that the speaker isn't properly ignoring, i.e., isn't actually properly ignoring, is ruled out by *S*'s evidence. Lewisian contextualism is meant to explain sceptical intuitions, so let's consider a particular sceptical intuition. Imagine a context where:

- I'm engaged in sceptical doubts;
- there is beer in the fridge
- I've forgotten what's in the fridge; and

- I've got normal vision, so if I check the fridge I'll see what's in it.

In that context it seems (10) is false, since it would only be true if Cartesian doubts weren't salient.

(10) If I were to look in the fridge and ignore Cartesian doubts, then I'd know there is beer in the fridge.

But the only way to get that to come out false, and false for the right reasons, is to fix on which worlds we're actually ignoring (i.e., include in the quantifier domain worlds where I'm the victim of an evil demon), but look at worlds that would be ruled out with the counterfactually available evidence. We don't want the sentence to be false because I've actually forgotten what's in the fridge. And we don't want it to be true because I would be ignoring Cartesian possibilities. In the terminology above, we would need *If p were true, all Fs would be Gs* to mean *If p were true, then every actual C that were F would also be G*. We haven't got any reason yet to think that's even a *possible* disambiguation of (10).

But let's make things easy for the contextualist and assume that it is. Stanley's point is that the contextualist needs even more than this. They need it to be by far the *preferred* disambiguation, since in the context I describe the natural reading of (10) (given sceptical intuitions) is that it is false because my looking in the fridge wouldn't rule out Cartesian doubts. And they need it to be the preferred reading even though there are alternative readings that are (a) easier to describe, (b) of a kind more commonly found, and (c) true. Every principle of contextual disambiguation we have pushes us away from thinking this is the preferred disambiguation. This is the deeper challenge Stanley raises for contextualists, and it hasn't yet been solved.

4.3   Knowledge By Indifference and By Wealth

Gillian Russell and John Doris (2009) argue that Jason Stanley's account of knowledge leads to some implausible attributions of knowledge. Insofar as my theory agrees with Stanley's about the kinds of cases they are worried about, their objections are also objections to my theory. I'm going to argue that Russell and Doris's objections turn on principles that are *prima facie* rather plausible, but which ultimately we can reject for independent reasons.[33]

Their objection relies on variants of the kind of case Stanley uses heavily in his (2005) to motivate a pragmatic constraint on knowledge. Stanley imagines a character who has evidence which would normally suffice for knowledge that p, but is faced with a decision where A is both the right thing to do if p is true, and will lead to a monumental material loss if p is false. Stanley intuits, and argues, that this is enough that they cease to know that p. I agree, at least as long as the gains from doing A are low enough that doing A amounts to a bet on p at insufficiently favourable odds to be reasonable in the agent's circumstance.

---

[33]I think the objections I make here are similar in spirit to those Stanley made in a comments thread on Certain Doubts, though the details are new. The thread is at http://el-prod.baylor.edu/certain_-doubts/?p=616

Russell and Doris imagine two kinds of variants on Stanley's case. In one variant the agent doesn't care about the material loss. As I'd put it, the agent's indifference to material odds shortens the odds of the bet. That's because costs and benefits of bets should be measured in something like utils, not something like dollars. As Russell and Doris put it, "you should have reservations ... about what makes [the knowledge claim] true: not giving a damn, however enviable in other respects, should not be knowledge-making." (Russell and Doris, 2009, 432). Their other variant involves an agent with so much money that the material loss is trifling to them. Again, this lowers the effective odds of the bet, so by my lights they may still know that p. But this is somewhat counter-intuitive. As Russell and Doris say, "[m]atters are now even dodgier for practical interest accounts, because *money* turns out to be knowledge making." (Russell and Doris, 2009, 433) And this isn't just because wealth can purchase knowledge. As they say, "money may buy the *instruments* of knowledge ... but here the connection between money and knowledge seems rather too direct." (Russell and Doris, 2009, 433)

The first thing to note about this case is that indifference and wealth aren't really producing knowledge. What they are doing is more like defeating a defeater. Remember that the agent in question had enough evidence, and enough confidence, that they would know p were it not for the practical circumstances. As I argued in the previous section, practical considerations enter debates about knowledge through two main channels: through the definition of belief, and through distinctive kinds of defeaters. It seems the second channel is particularly relevant here. And we have, somewhat surprisingly, independent evidence to think that indifference and wealth do matter to defeaters.

Consider two variants on Gilbert Harman's 'dead dictator' example (Harman, 1973, 75). In the original example, an agent reads that the dictator has died through an actually reliable source. But there are many other news sources around, defeaters, such that if the agent read them, she would lose her belief.

In the first variant, the agent simply does not care about politics. It's true that there are many other news sources around that are ready to mislead her about the dictator's demise. But she has no interest in looking them up, nor is she at all likely to look them up. She mostly cares about sports, and will spend most of her day reading about baseball. In this case, the misleading news sources are too distant, in a sense, to be defeaters. So she still knows the dictator has died. Her indifference towards politics doesn't generate knowledge - the original reliable report is the knowledge generator - but her indifference means that a would-be defeater doesn't gain traction.

In the second variant, the agent cares deeply about politics, and has masses of wealth at hand to ensure that she knows a lot about it. Were she to read the misleading reports that the dictator has survived, then she would simply use some of the very expensive sources she has to get more reliable reports. Again this suffices for the misleading reports not to be defeaters. Even before the rich agent exercises her wealth, the fact that her wealth gives her access to reports that will correct for misleading reports means that the misleading reports are not actually defeaters. So with her wealth she knows things she wouldn't otherwise know, even before her money goes to work. Again, her money doesn't generate knowledge – the original reliable

report is the knowledge generator – but her wealth means that a would-be defeater doesn't gain traction.

The same thing is true in Russell and Doris's examples. The agent has quite a bit of evidence that p. That's why she knows p. There's a potential practical defeater for p. But due to either indifference or wealth, the defeater is immunised. Surprisingly perhaps, indifference and/or wealth can be the difference between knowledge and ignorance. But that's not because they can be in any interesting sense 'knowledge makers', any more than I can make a bowl of soup by preventing someone from tossing it out. Rather, they can be things that block defeaters, both when the defeaters are the kind Stanley talks about, and when they are more familiar kinds of defeaters.

4.4   Stakes, Odds and Experiments

In a so far unpublished note, Mark Schroeder (2010) has argued that interest-relative invariantists have erred by stressing variation in stakes as being relevant to knowledge. He argues, using examples of forced choice, that what is really relevant is the odds at which the agent has to make bets. Of course due to the declining marginal utility of material goods, high stakes bets will often be long odds bets. So there's a correlation between stakes and odds. But when the correlation comes apart, Schroeder argues convincingly that it's the odds and not the stakes that are relevant to knowledge.

The view of belief I've been defending agrees with Schroeder's judgments. Interests affect belief because whether someone believes p depends *inter alia* on whether their credence in p is high enough that any bet on p they actually face is a good bet. Raising the stakes of any bet on p does not change that, but changing the odds of the bets on p they face does change it. And that explains why agents don't have knowledge, or even justified belief, in some of the examples that motivate other interest-relative invariantists.

Although I think my view gets those cases right, I don't take those examples to be a crucial part of the argument for the view. The core argument I offered in Weatherson (2005) is that the view provides a better answer to the challenge of how we should integrate credences and beliefs into a single model. If it turned out that the facts about the examples were less clear than we thought, that wouldn't *undermine* the argument for my view, since those facts weren't part of the original argument. But if it turned out that the facts about those examples were quite different to what my view about belief and knowledge predicts, that may *rebut* the view, since it would then be shown to make false predictions.

This kind of rebuttal may be suggested by various recent experimental results, such as the results in May et al. (forthcoming) and Feltz and Zarpentine (forthcoming). I'm going to concentrate on the latter set of results here, though I think that what I say will generalise to related experimental work.[34] Feltz and Zarpentine gave subjects related vignettes, such as the following pair. (Each subject only received one of the pair.)

---

[34]Note to editors: Because this work is not yet in press, I don't have page numbers for any of the quotes from Feltz and Zarpentine. That should be fixed by the time *we* are ready for press.

High Stakes Bridge John is driving a truck along a dirt road in a caravan of trucks. He comes across what looks like a rickety wooden bridge over a yawning thousand foot drop. He radios ahead to find out whether other trucks have made it safely over. He is told that all 15 trucks in the caravan made it over without a problem. John reasons that if they made it over, he will make it over as well. So, he thinks to himself, 'I know that my truck will make it across the bridge.'

Low Stakes Bridge John is driving a truck along a dirt road in a caravan of trucks. He comes across what looks like a rickety wooden bridge over a three foot ditch. He radios ahead to find out whether other trucks have made it safely over. He is told that all 15 trucks in the caravan made it over without a problem. John reasons that if they made it over, he will make it over as well. So, he thinks to himself, 'I know that my truck will make it across the bridge.' (Feltz and Zarpentine, forthcoming, ??)

Subjects were asked to evaluate John's thought. And the result was that 27% of the participants said that John does not know that the truck will make it across in **Low Stakes Bridge**, while 36% said he did not know this in **High Stakes Bridge**. Feltz and Zarpentine say that these results should be bad for interest-relativity views. But it is hard to see just why this is so.

Note that the change in the judgments between the cases goes in the direction that my view predicts. The change isn't trivial, even if due to the smallish sample size it isn't statistically significant in this sample. But should a view like mine have predicted a larger change? To figure this out, we need to ask three questions.

1. What are the costs of the bridge collapsing in the two cases?
2. What are the costs of not taking the bet, i.e., not driving across the bridge?
3. What is the rational credence to have in the bridge's sturdiness given the evidence John has?

None of these are specified in the story given to subjects, so we have to guess a little as to what the subjects' views would be.

Feltz and Zarpentine say that the costs in "High Stakes Bridge [are] very costly—certain death—whereas the costs in Low Stakes Bridge are likely some minor injuries and embarrassment." (Feltz and Zarpentine, forthcoming, ??) I suspect both of those claims are wrong, or at least not universally believed. A lot more people survive bridge collapses than you may expect, even collapses from a great height.[35] And once the road below a truck collapses, all sorts of things can go wrong, even if the next bit of ground is only 3 feet away. (For instance, if the bridge collapses unevenly, the truck could roll, and the driver would probably suffer more than minor injuries.)

We aren't given any information as to the costs of not crossing the bridge. But given that 15 other trucks, with less evidence than John, have decided to cross the

---

[35]In the West Gate bridge collapse in Melbourne in 1971, a large number of the victims were underneath the bridge; the people on top of the bridge had a non-trivial chance of survival. That bridge was 200 feet above the water, not 1000, but I'm not sure the extra height would matter greatly. Again from a slightly lower height, over 90% of people on the bridge survived the I-35W collapse in Minneapolis in 2007.

bridge, it seems plausible to think they are substantial. If there was an easy way to avoid the bridge, presumably the *first* truck would have taken it.

But the big issue is the third question. John has a lot of information that the bridge will support his truck. If I've tested something for sturdiness two or three times, and it has worked, I won't even think about testing it again. Consider what evidence you need before you'll happily stand on a particular chair to reach something in the kitchen, or put a heavy television on a stand. Supporting a weight is the kind of thing that either fails the first time, or works fairly reliably. Obviously there could be some strain-induced effects that cause a subsequent failure[36], but John really has a lot of evidence that the bridge will support him.

Given those three answers, it seems to me that it is a reasonable bet to cross the bridge. At the very least, it's no more of an unreasonable bet than the bet I make every day crossing a busy highway by foot. So I'm not surprised that 64% of the subjects agreed that John knew the bridge would hold him. At the very least, that result is perfectly consistent with the my views about belief and knowledge, if we make plausible assumptions about how the subjects would answer the three numbered questions above.

And as I've stressed, these experiments are only a problem for my view if the subjects are reliable. I can think of two reasons why they might not be. First, subjects tend to massively discount the costs and likelihoods of traffic related injuries. In most of the country, the risk of death or serious injury through motor vehicle accident is much higher than the risk of death or serious injury through some kind of crime or other attack, yet most people do much less to prevent vehicles harming them than they do to prevent criminals or other attackers harming them.[37] Second, only 73% of this subjects in *this very experiment* said that John knows the bridge will support him in **Low Stakes Bridge**. This is just absurd. Unless the subjects endorse an implausible kind of scepticism, something has gone wrong with the experimental design. Given the fact that the experiment points broadly in the direction of the theory I favour, and that with some plausible assumptions it is perfectly consistent with that theory, and that the subjects seem unreasonably sceptical to the point of unreliability about epistemology, I don't think this kind of experimental work threatens my view.

---

[36] As I believe was the case in the I-35W collapse.

[37] See the massive drop in the numbers of students walking or biking to school, reported in Ham et al. (2008), for a sense of how big an issue this is.

Bibliography

Blome-Tillmann, Michael. 2009a. "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.

—. 2009b. "Knowledge and Presuppositions." *Mind* 118:241–294, doi:10.1093/mind/fzp032.

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

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

Fantl, Jeremy and McGrath, Matthew. 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.

Feltz, Adam and Zarpentine, Chris. forthcoming. "Do You Know More When It Matters Less?" *Philosophical Psychology* Retrieved from http://faculty.schreiner.edu/adfeltz/Papers/Know%20more.pdf.

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

Ham, Sandra A., Martin, Sarah, and Kohl III, Harold W. 2008. "Changes in the percentage of students who walk or bike to school-United States, 1969 and 2001." *Journal of Physical Activity and Health* 5:205–215.

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

Harman, Gilbert. 1973. *Thought*. Princeton: Princeton University Press.

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

Jackson, Frank. 1991. "Decision Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101:461–82.

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

—. 1996b. "Philosophy and Commonsense: The Case of Weakness of Will." In Michaelis Michael and John O'Leary-Hawthorne (eds.), *The Place of Philosophy in the Study of Mind*, 141–157. Norwell, MA: Kluwer, doi:10.1017/CBO9780511606977.005.

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

—. 1982. "Logic for Equivocators." *Noûs* 16:431–441. Reprinted in *Papers in Philosophical Logic*, pp. 97-110.

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

Lipsey, R. G. and Lancaster, Kelvin. 1956-1957. "The General Theory of Second Best." *Review of Economic Studies* 24:11–32, doi:10.2307/2296233.

Maher, Patrick. 1993. *Betting on Theories*. Cambridge: Cambridge University Press.

May, Joshua, Sinnott-Armstrong, Walter, Hull, Jay G., and Zimmerman, Aaron. forthcoming. "Practical Interests, Relevant Alternatives, and Knowledge Attributions: an Empirical Study." *Review of Philosophy and Psychology* , doi:10.1007/s13164-009-0014-3.

Neta, Ram. 2007. "Anti-intellectualism and the Knowledge-Action Principle." *Philosophy and Phenomenological Research* 75:180–187, doi:10.1111/j.1933-1592.2007.00069.x.

Ramsey, Frank. 1926. "Truth and Probability." In D. H. Mellor (ed.), *Philosophical Papers*, 52–94. Cambridge: Cambridge University Press.

Russell, Gillian and Doris, John M. 2009. "Knowledge by Indifference." *Australasian Journal of Philosophy* 86:429–437, doi:10.1080/00048400802001996.

Schroeder, Mark. 2010. "Stakes, Withholding and the Subject-Sensitivity of 'Knows'." Retrieved from http://www-bcf.usc.edu/~maschroe/research/Schroeder_Stakes_and_Withholding.pdf.

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

Stanley, Jason and Szabó, Zoltán Gendler. 2000. "On Quantifier Domain Restriction." *Mind and Language* 15:219–61, doi:10.1111/1468-0017.00130.

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

Turri, John. 2010. "On the Relationship between Propositional and Doxastic Justification." *Philosophy and Phenomenological Research* 80:312–326, doi:10.1111/j.1933-1592.2010.00331.x.

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

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

—. 2005. "Can We Do Without Pragmatic Encroachment?" *Philosophical Perspectives* 19:417–443, doi:10.1111/j.1520-8583.2005.00068.x.

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

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

—. 2000. *Knowledge and its Limits*. Oxford University Press.