Defending Interest-Relative Invariantism

Blinded for review

In recent years a number of authors have defended the interest-relativity of various epistemological claims, such as claims of the form *S knows that p*, or *S has a justified belief that p*. Views of this form are floated by John Hawthorne (2004), and endorsed by Jeremy Fantl and Matthew McGrath (2002; 2009), Jason Stanley Stanley (2005) and Brian Weatherson (2005). The various authors differ quite a lot in how much interest-relativity they allow, and even more in their purported explanations for the interest-relativity, but what is common is the endorsement of some kind of interest-relativity in statements like *S knows that p*, or *S has a justified belief that p*.

These views have, quite naturally, drawn a range of criticisms. The primary purpose of this paper is to respond to these criticisms and, as it says on the tin, defend interest-relative invariantism, or IRI for short.[1] But to defend the view, I first need to clarify three features of the view. The best version of IRI, and the only one I'm interested in defending, has these three features:

- Odds, not stakes, are primarily what matter to knowledge.
- Interests create defeaters.
- Interest-relativity is an existential claim; it says that interests sometimes matter, not that they always do.

In the first section, I'll say more about each of these three points. In later sections, I'll apply these points to defend IRI against criticisms.

1 Clarifying IRI

1.1 Odds and Stakes

It is common to describe IRI as a theory where in 'high stakes' situations, more evidence is needed for knowledge than in 'low stakes' situations. But this is at best misleading. What really matters are the odds on any bet-like decision the agent faces with respect to the target proposition. More precisely, 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.

---

[1]'Interest-relative invariantism' is Jason Stanley's term for the view that 'knows' is not context-sensitive, but whether S knows that p might depend on S's interests. Some critics, such as Michael Blome-Tillmann (2009), call IRI '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 the critics target certainly isn't defined in opposition to *this* generalisation.

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. Now in practice due to the declining marginal utility of material goods, high stakes situations will usually be situations where an agent faces long odds. But it is the odds that matter to knowledge, not the stakes.[2]

Some confusion on this point may have been caused by the Bank Cases that Stanley uses, and the Train Cases that Fantl and McGrath use, to motivate IRI. In those cases, the authors lengthen the odds the relevant agents face by increasing the potential losses the agent faces by getting the bet wrong. But we can make the same point by decreasing the amount the agent stands to gain by taking the bet. Let's go through a pair of cases, which I'll call the Map Cases, that illustrate this.

High Cost Map: Zeno is walking to the Mysterious Bookshop in lower Manhattan. He's pretty confident that it's on the corner of Warren Street and West Broadway. But he's been confused about this in the past, forgetting whether the east-west street is Warren or Murray, and whether the north-south street is Greenwich, West Broadway or Church. In fact he's right about the location this time, but he isn't justified in having a credence in his being correct greater than about 0.95. While he's walking there, he has two options. He could walk to where he thinks the shop is, and if it's not there walk around for a few minutes to the nearby corners to find where it is. Or he could call up directory assistance, pay \$1, and be told where the shop is. Since he's confident he knows where the shop is, and there's little cost to spending a few minutes walking around if he's wrong, he doesn't do this, and walks directly to the shop.

Low Cost Map: Just like the previous case, except that Zeno has a new phone with more options. In particular, his new phone has a searchable map, so with a few clicks on the phone he can find where the store is. Using the phone has some very small costs. For example, it distracts him a little, which marginally raises the cost of bumping into another pedestrian. But the cost is very small compared to the cost of getting the location wrong. So even though he is very confident that he knows where the shop is, he double checks while walking there.

I think the Map Cases are like the Bank Cases, Train Cases etc in all important respects. I think Zeno knows where the shop is in High Cost Map, and doesn't know in Low Cost Map. And he doesn't know in Low Cost Map because the location of the shop has suddenly become the subject matter of a bet at very long

---

[2]Similar points to these are developed in a so far unpublished note by Mark Schroeder (2010).

odds. You should think of Zeno's not checking the location of the shop on his phone-map as a bet on the location of the shop. If he wins the bet, he wins a few seconds of undistracted strolling. If he loses, he has to walk around a few blocks looking for a store. The disutility of the loss seems easily twenty times greater than the utility of the gain, and by hypothesis the probability of winning the bet is no greater than 0.95. So he shouldn't take the bet. Yet if he knew where the store was, he would be justified in taking the bet. So he doesn't know where the store is. Now this is not a case where higher *stakes* defeat knowledge. If anything, the stakes are lower in Low Cost Map. But the relevant odds are longer, and that's what matters to knowledge.[3]

1.2 Interests are Defeaters

Interests have a somewhat roundabout effect on knowledge. The IRI story goes something like this. If the agent has good but not completely compelling evidence for p, that is sometimes but not always sufficient for knowledge that p if everything else goes right. It isn't sufficient if they face a choice where the right thing to do is different to the right thing to do conditional on p. That is, if adding p to their background information would make a genuinely bad choice look like a good choice, they don't know that p. The facts about this choice, both the objective facts about it and the agent's beliefs about it, can defeat the agent's claim to knowledge that p.

Cases where knowledge is defeated because if the agent did know p, that would lead to problems elsewhere in their cognitive system, have a few quirky features. In particular, whether the agent knows p can depend on very distant features. Consider the following kind of case.

> **Confused Student**
>
> Con is systematically disposed to affirm the consequent. That is, if he notices that he believes both p and $q \rightarrow p$, he's disposed to either infer q, or if that's impermissible given his evidence, to ditch his belief in the conjunction of p and $q \rightarrow p$. Con has completely compelling evidence for both $q \rightarrow p$ and $\neg q$. He has good but less compelling evidence for p. And this evidence tracks the truth of p in

---

[3]Note that I'm not claiming that it is intuitive that Zeno has knowledge in High Cost Map, but not that Low Cost Map. Nor am I claiming that we should believe IRI because it gets the Map Cases right. In fact, I don't believe either of those things. In fact, I believe Zeno has knowledge in High Cost Map and not in Low Cost Map because I believe IRI is correct, and that's what IRI says about the case. It is sometimes assumed, e.g, in the experimental papers I'll discuss below, that pairs of cases like these are meant to *motivate*, and not just *illustrate* IRI. I can't speak for everyone's motivations, but I'm only using these cases as illustrations, not motivations.

just the right way for knowledge. On the basis of this evidence, Con believes p. Con has not noticed that he believes both p and $q \rightarrow p$. If he did, he's unhesitatingly drop his belief that p, since he'd realise the alternatives (given his dispositions) involved dropping belief in a compelling proposition. Two questions:

- Does Con know that p?
- If Con were to think about the logic of conditionals, and reason himself out of the disposition to affirm the consequent, would he know that p?

I think the answer to the first question is *No*, and the answer to the second question is *Yes*. As it stands, Con's disposition to affirm the consequent is a doxastic defeater of his putative knowledge that p. Put another way, p doesn't cohere well enough with the rest of Con's views for his belief that p to count as knowledge. To be sure, p coheres well enough with those beliefs by objective standards, but it doesn't cohere at all by Con's lights. Until he changes those lights, it doesn't cohere well enough to be knowledge.

I don't expect exactly everyone will agree with those judgments. Some people will simply reject that this kind of coherence by one's own lights is necessary for knowledge. Others might even reject the whole idea of doxastic defeaters. But I think the picture I've just sketched, one which puts reasonably tight coherence constraints on knowledge, is plausible enough to use in a defence of IRI. It certainly isn't so implausible that committing to it amounts to a *reductio* of one's views. Yet as we'll frequently see below, some criticisms of IRI do suggest that any theory that allows for these kinds of coherence constraints or doxastic defeaters is thereby shown to be false. I'm going to take that suggestion to be a *reductio* of the criticisms.

1.3 IRI is an Existential Theory

IRI theorists do not typically say that interests are *always* relevant to knowledge. In fact, they hardly could be. If p is not true, or the agent has very little evidence for p, the agent does not know p whatever their interests. But an assumption that seems to shared by both some critics and some proponents of IRI is that IRI rests on some universal epistemic principles, not just on various existential principles. For instance, Jeremy Fantl and Matthew McGrath use a lot of principles like (JJ) in deriving IRI.

(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)

Now we it turns out, I think (JJ) is false. (I think it fails in cases where the agent is seriously mistaken about the risks and payoffs involved in doing φ, for instance.) But more importantly, we don't need anything nearly as strong as this to derive. As long as there is *some* sufficiently large range of cases where (JJ) holds, we'll be able to establish the existence of *some* pair of cases which differ in whether the agent knows that p in virtue of the interests the agent has.

Relatedly, the argument for a version of IRI in Weatherson (2005) makes frequent appeal to standard Bayesian decision theory. This might suggest that such an argument stands and falls with the success of consequentialism in decision theory. (I mean to use 'consequentialism' here roughly in the sense that it is used in Hammond (1988).) But again, this suggestion would be false. If consequentialism is true in some range of cases, we'll be able to use similar techniques to the ones used in that paper to show that there are the kinds of pairs of cases that IRI say exist.

2 Experimental Objections

As I mentioned in the discussion of the Map Cases, I don't think the argument for IRI rests on judgments, or intuitions, about similar cases. Rather, IRI can be independently motivated, and is so motivated in the books and papers cited in the first paragraph of this paper. It's a happy result, in my view, that IRI gets various Bank Cases and Map Cases right, but not essential to the view. If it turned out that the facts about the examples were less clear than we thought, that wouldn't *undermine* the argument for IRI, since those facts weren't part of the best arguments for IRI. But if it turned out that the facts about those examples were quite different to what IRI 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.[4] In fact, I think the experiments don't really tell against IRI, because IRI doesn't make *any* unambiguous predictions about the cases which could possibly be experimentally refuted. The reason for this is related to the first point made in section one: it is odds, not stakes, that are most important.

Feltz and Zarpentine gave subjects related vignettes, such as the following pair. (Each subject only received one of the pair.)

---

[4]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.

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 IRI seems to predict. 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 IRI 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?

IRI predicts that there is knowledge in Low Stakes Bridge but not in High Stakes Bridge only if the following equation is true:

$$\frac{C_H}{G + C_H} > x > \frac{C_L}{G + C_L}$$

, where G is the gain the driver gets from taking a non-collapsing bridge rather than driving around (or whatever the alternative is), C_H is the cost of being on a collapsing bridge in High Stakes Bridge, C_L is the cost of being on a collapsing bridge in Low Stakes Bridge, and x is the probability that the bridge will collapse.

I assume x is constant between the two cases. If that equation holds, then taking the bridge, i.e., acting as if the bridge is safe, maximises expected utility in Low Stakes Bridge but not High Stakes Bridge. So in High Stakes Bridge, adding the proposition that the bridge won't collapse to the agent's cognitive system produces incoherence, since the agent won't (at least rationally) act as if the bridge won't collapse. So if the equation holds, the agent's interests in avoiding C_H creates a doxastic defeater in High Stakes Bridge.

But does the equation hold? Or, more relevantly, did the subjects of the experiment believe that the equation hold. None of the four variables has their values clearly entailed by the story, 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.[5] 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 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. If G is large enough, and C_H small enough, then the only way for this equation to hold will be for x to be low enough that we'd have independent reason to say that the driver doesn't know the bridge will hold.

But what is the value of x? 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

---

[5]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.

could be some strain-induced effects that cause a subsequent failure[6], 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 IRI, 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 IRI 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.[7] 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 IRI, 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 IRI.

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 version of IRI 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.[8]

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,

---

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

[7]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.

[8]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

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.

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 first section, practical considerations enter debates about knowledge in part because they are distinctive kinds of defeaters. It seems that's what is going on 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   Temporal Embeddings

Michael Blome-Tillmann (2009) has argued that tense-shifted knowledge ascriptions can be used to show that his version of Lewisian contextualism is preferable to IRI. His argument uses a variant of the well-known bank cases.[9] 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 (1).

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

---

[9]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.

But I'm not sure why this case should be problematic for any version of IRI, and very unsure why it should even look like a *reductip* of IRI. 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 (1) true if IRI is true. Now actually I agree that this change in perception could be enough to make (1) 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 IRI.

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 IRI. 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 (1) 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.

What if Hannah is, on Thursday, irrationally ignoring ¬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. But it's worth remembering that if Hannah does irrationally ignore ¬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's what we saw in Con's case in section 1, and it's all we see here as well. Note also that Con will know p after fixing his consequent-affirming disposition but not before. That's just what happens with Hannah; a distant change in her cognitive system will remove a defeater, and after it does she gets more knowledge.

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. But once we think through Hannah's case, we see perfectly good reasons, independent of IRI, to endorse IRI's prediction about the case.

5 Problematic Conjunctions

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. Clearly, (1) and (2) are true.

(1)   Ringo has enough money to buy the computer.

(2)   Ringo can afford the computer.

And (3) is true as well, though there's at least a reading of (4) where it is false.

(3)   George has enough money to buy the computer.

(4)   George can afford the computer.

Focus for now on (3). It is a bad idea for George to buy the computer; he won't be able to pay his rent. But he has enough money to do so; the computer costs $2000, and he has $6000 in the bank. So (3) is true. Admittedly there are things close to (3) that aren't true. He hasn't got enough money to buy the computer and pay his rent. You might say that he hasn't got enough money to buy the computer given his other financial obligations. But none of this undermines (3). The point of this little story is to respond to another argument Blome-Tillmann offers against IRI. Here is how he puts the argument. (Again I've changed the numbering and some terminology for consistency with this paper.)

> Suppose that John and Paul have exactly the same evidence, while John is in a low-stakes situation towards p and Paul in a high-stakes situation towards p. Bearing in mind that IRI is the view that whether one knows p depends on one's practical situation, IRI entails that one can truly assert:
>
> (5)   John and Paul have exactly the same evidence for p, but only John has enough evidence to know p, Paul doesn't.

And this is meant to be a problem, because (5) is intuitively false.

But IRI doesn't entail any such thing. Paul does have enough evidence to know that p, just like George has enough money to buy the computer. Paul can't know that p, just like George can't buy the computer, because of their practical situations. But that doesn't mean he doesn't have enough evidence to know it. So, contra Blome-Tillmann, IRI doesn't entail this problematic conjunction.

In a footnote attached to this, Blome-Tillmann tries to reformulate the argument.

> I take it that having enough evidence to 'know p' in C just means having evidence such that one is in a position to 'know p' in C, rather than having evidence such that one 'knows p'. Thus, another way to formulate (5) would be as follows: 'John and Paul have exactly the same evidence for p, but only John is in a position to know p, Paul isn't.'

The 'reformulation' is obviously bad, since having enough evidence to know p isn't the same as being in a position to know it, any more than having enough money to buy the computer puts George in a position to buy it. But might there be a different problem for IRI here? Might it be that IRI entails (6), which is false?

(6)   John and Paul have exactly the same evidence for p, but only John is in a position to know p, Paul isn't.

There isn't a problem with (6) because almost any epistemological theory will imply that conjunctions like that are true. In particular, any epistemological theory that allows for the existence of defeaters to not supervene on the possession of evidence will imply that conjunctions like (6) are true. Again, it matters a lot that IRI is suggesting that traditional epistemologists did not notice that there are distinctively pragmatic defeaters. Once we see that, we'll see that conjunctions like (6) are not surprising at all.

Consider again Con, and his friend Mod who is disposed to reason by modus ponens and not by affirming the consequent. We could say that Con and Mod have the same evidence for p, but only Mod is in a position to know p. There are only two ways to deny that conjunction. One is to interpret 'position to know' so broadly that Con is in a position to know p because he could change his inferential dispositions. But then we might as well say that Paul is in a position to know p because he could get into a different 'stakes' situation. Alternatively, we could say that Con's inferential dispositions count as a kind of evidence against p. But that stretches the notion of evidence beyond a breaking point. Note that we didn't say Con had any *reason* to affirm the consequent, just that he does. Someone might adopt, or change, a poor inferential habit because they get new evidence. But they need not do so, and we shouldn't count their inferential habits as evidence they have.

If that case is not convincing, we can make the same point with a simple Gettier-style case.

Getting the Job In world 1, at a particular workplace, someone is about to be promoted. Agnetha knows that Benny is the management's favourite choice for the promotion. And she also knows that Benny is Swedish. So she comes to believe that the promotion will go to someone Swedish. Unsurprisingly, management does choose Benny, so Agnetha's belief is true.

World 2 is similar, except there it is Anni-Frid who knows that Benny is the management's favourite choice for the promotion, that Benny is Swedish. So *she* comes to believe that the promotion will go to someone Swedish. But in this world Benny quits the workplace just before the promotion is announced, and the management unexpectedly passes over a lot of Danish workers to promote another Swede, namely Bj orn. So Anni-Frid's belief that the promotion will go to someone Swedish is true, but not in a way that she could have expected.

In that story, I think it is clear that Agnetha and Anni-Frid have exactly the same evidence that the job will go to someone Swedish, but only Agnetha is in a position to know this, Anni-Frid is not. The fact that an intermediate step is false in Anni-Frid's reasoning, but not Agnetha's, means that Anni-Frid's putative knowledge is defeated, but Agnetha's is not. And when that happens, we can have differences in knowledge without differences in evidence. So it isn't an argument against IRI that it allows differences in knowledge without differences in evidence.

6   Non-Consequentialist Casees

None of the replies yet have leaned heavily on the third point from section one, the fact that IRI is an existential claim. This reply will make heavy use of that fact.

If an 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 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.[10]

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

---

[10]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.

true, is 0, and that's higher than the conditional expected value of any other act, conditional on p. If you thought that, you'd certainly be in agreement with the orthodox decision-theoretic treatment of this problem.

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

My reasons for this go back to Jackson's original discussion of the puzzle. Let the agent be a doctor, the actions 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.[11]

> 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)

---

[11]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. But it's important to see how IRI handles the cases that Brown discusses, since these cases are among the strongest challenges that have been raised to IRI.

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 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, that the surgeon should check the notes, even if she believes *and knows* that the left kidney is the diseased one.

There is a very general point here. The best arguments for IRI start with the role that knowledge plays in a particular theory of decision or reasoning.

References

Blome-Tillmann, Michael. 2009. "Contextualism, Subject-Sensitive Invariantism, and the Interaction of 'Knowledge'-Ascriptions with Modal and Temporal Operators." *Philosophy and Phenomenological Research* 79:315–331, doi:10.1111/j.1933-1592.2009.00280.x.

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. 2002. "Evidence, Pragmatics, and Justification." *Philosophical Review* 111:67–94, doi:10.1215/00318108-111-1-67.

—. 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.

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.

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.

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.

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