

# Defending Interest-Relative Invariantism

Brian Weatherson

---

In recent years a number of authors have defended the interest-relativity of knowledge and justification. Views of this form are floated by John Hawthorne (2004), and endorsed by Jeremy Fantl and Matthew McGrath (2002; 2009), Jason Stanley (2005) and Brian Weatherson (2005). The various authors differ quite a lot in how much interest-relativity they allow, but what is common is the defence of interest-relativity.

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. But I don't plan to defend every possible version of IRI, only a particular one. Most of the critics of IRI have assumed that it must have some or all of the following features.

1. It is harder to know things in high-stakes situations than in low-stakes situations.
2. There is an interest-sensitive constituent of knowledge.
3. IRI stands and falls with some principles connecting knowledge and action, such as the principles found in Hawthorne and Stanley (2008).

My preferred version of IRI has none of these three features.<sup>1</sup>

First, it says that knowledge changes when the **odds** an agent faces change, not when the **stakes** change. 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. And interests affect knowledge largely because they affect belief. Raising the stakes of any bet on  $p$  does not directly change whether an agent believes  $p$ , but changing the odds of the bets on  $p$  they face does change it. In practice raising the stakes changes the odds due to the declining marginal utility of material goods. So in practice high-stakes situations are typically long-odds situations. But knowledge is hard in those situations because they are long-odds situations, not because they are high-stakes situations.

So my version of IRI says that knowledge differs between these two cases.

**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 likelihood 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 about where the shop is, he double checks while walking there.

I think the Map Cases are like the various cases that have been used to motivate interest-relativity<sup>2</sup> 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 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.

Second, on this version of IRI, interests matter because there are interest-sensitive defeaters, not because interests form any kind of new condition on knowledge, alongside truth, justification, belief and so on. In particular, interests matter because there are interest-relative coherence constraints on knowledge. Some coherence constraints, I claim, are not interest-relative. If an agent believes  $\neg p$ , that belief defeats her purported knowledge that  $p$ , even if the belief that  $p$  is true, justified, safe, sensitive and so on. It is tempting to try to posit a further coherence condition.

**Practical Coherence** An agent does not know that  $p$  if she prefers  $\varphi$  to  $\psi$  unconditionally, but prefers  $\psi$  to  $\varphi$  conditional on  $p$ .

But that is too strong. For reasons similar to those gone over at the start of Hawthorne (2004), it would mean we know nearly nothing. A more plausible condition is:

**Relevant Practical Coherence** An agent does not know that  $p$  if she prefers  $\varphi$  to  $\psi$  unconditionally, but prefers  $\psi$  to  $\varphi$  conditional on  $p$ , for any  $\varphi, \psi$  that are relevant given her interests.

When this condition is violated, the agent's claim to knowledge is defeated. As we'll see below, defeaters behave rather differently to constituents of knowledge. Some things which could not plausibly be grounds for knowledge could be defeaters to defeaters for knowledge.

**Relevant Practical Coherence** suffices, at least among agents who are trying to maximise expected value, to generate an interest-relativity to knowledge. The general structure of the case should be familiar from the existing literature. Let  $p$  be a proposition that is true, believed by the agent, and strongly but not quite conclusively supported by their evidence. Let  $B$  be a bet that has a small positive return if  $p$ , and a huge negative return if  $\neg p$ . Assume the agent is now offered the bet, and let  $\varphi$  be declining the bet, and  $\psi$  be accepting the bet. Conditional on  $p$ , the bet wins, so the agent prefers the small positive payout, so prefers  $\psi$  to  $\varphi$  conditional on  $p$ . But the bet has a massively negative expected return, so unconditionally the agent does not want it. That is, unconditionally she prefers  $\varphi$  to  $\psi$ . Once the bet is offered, the actions  $\varphi$  and  $\psi$  become relevant given her interests, so by **Relevant Practical Coherence** she no longer knows  $p$ . So for such an agent, knowledge is interest-relative.

This suggests our third point. This version of IRI does not take IRI to be a consequence of more general principles about knowledge and action. It simply says that there exist at least one pair of cases where the only relevant difference between agents in the two cases concerns their interests, but one knows that  $p$  and the other does not.<sup>3</sup> I happen to think that most of the general principles that philosophers have used to try to derive IRI are false. But since IRI is much weaker than those principles, that is no reason to conclude IRI is false.<sup>4</sup>

The existence of interest-relativity is then quite a weak claim. There are plenty of stronger claims in the area we could make. I prefer, for instance, a version of IRI where being offered bets like  $B$  defeats knowledge

that  $p$  even if the agent does not have the preferences I ascribed above. (That could be because she isn't trying to maximise expected value, or because she's messed up the expected value calculations.) But knowledge could be interest-relative even if I'm wrong about those cases.

So I've set out a version of IRI that lacks three features often attributed to IRI. I haven't argued for that theory here - I do that at much greater length in (Author Paper 1). But I hope I've done enough to convince you that the theory is both a version of IRI, and not obviously false. In what follows, I'll argue that the theory is immune to the various challenges to IRI that have been put forward in the literature. This immunity is, I think, a strong reason to prefer this version of IRI.

## 1 Experimental Objections

I don't place as much weight as some philosophers do on the correlation between the verdicts of an epistemological theory and the gut reactions that non-experts have to tricky cases. And I don't think the best cases for IRI relies on such a correlation holding. The best case for IRI is that it integrates nicely with an independently supported theory of belief, and that it lets us keep a number of plausible principles without drifting into skepticism.<sup>5</sup> But still, it is nice to not have one's theory saying exorbitantly counterintuitive things. Various experimental results, such as the results in May et al. (forthcoming) and Feltz and Zarpentine (2010), might be thought to suggest that IRI does have consequences which are counterintuitive, or which at least run counter to the intuitions of some experimental subjects. 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. In fact, I think the experiments don't really tell against IRI, because IRI, at least in my preferred version, doesn't make *any* unambiguous predictions about the cases at the centre of the experiments. The reason for this is related to my insistence that we concentrate on the odds an agent faces, not the stakes she faces.

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

**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, 2010, 696)

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?

Conditional on the bridge not collapsing, the drivers presumably prefer taking the bridge to not taking it. And the actions of taking the bridge or going around the long way are relevant. So by **Relevant Practical Coherence**, the drivers know the bridge will not collapse in Low Stakes Bridge but not High Stakes Bridge if the following equation is true. (I assume all the other conditions for knowledge are met, and that there are no other salient instances of Relevant Practical Coherence to consider.)

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

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 won't collapse, 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, 2010, 702) 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.<sup>6</sup> 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 could be some strain-induced effects that cause a subsequent failure<sup>7</sup>, 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.<sup>8</sup> Second, only 73% of these subjects in *this very experiment* said that John knows the bridge will support him in **Low Stakes Bridge**. This is rather striking. Unless the subjects endorse an implausible kind of scepticism, something has gone wrong with the experimental design. But if the subjects are implausibly sceptical, then we shouldn't require our epistemological theory to track their gut reactions. (And if something has gone wrong with the experimental design, then obviously can't be used as the basis for any objection.) So 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.

## 2 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, and if successful their objections would generalise to other forms of IRI. 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.<sup>9</sup>

Their objection relies on variants of the kind of case Stanley uses heavily in his (2005) to motivate a pragmatic constraint on knowledge. Stanley considers the kinds of cases we used to derive IRI from **Relevant Practical Coherence**. So imagine an agent who faces a choice between accepting the status quo, call that  $\varphi$ , and taking some giant risk, call that  $\psi$ . The giant risk in this case will involve a huge monetary loss if  $\neg p$ , and a small non-monetary gain if  $p$ . Stanley says, and I agree, that in such a case the agent doesn't know  $p$ , even if their belief in  $p$  is true, well supported by evidence, and so on. Moreover, he says, had  $\psi$  not been a relevant option, the agent could have known  $p$ . I agree, and I think **Relevant Practical Coherence** explains these intuitions well.

Russell and Doris imagine two kinds of variants on Stanley's case. In one variant the agent doesn't care about the material loss associated with  $\psi \wedge \neg p$ . As I would put it, although their material wealth would decline precipitously in that case, their utility would not, because their utility is not tightly correlated with material wellbeing. Given that, the agent may well prefer  $\psi$  to  $\varphi$  unconditionally, and so would still know  $p$ . Russell and Doris don't claim this is a problem in itself, but they do think the conjunction of this with the previous paragraph is a problem. As they 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. Since the difference in utility between having, say, eight billion dollars and seven billion dollars is not that high, perhaps they will again prefer  $\psi$  to  $\varphi$  unconditionally, so still know  $p$ . But it is, allegedly, counterintuitive to have the knowledge that  $p$  turn on the agent's wealth. 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 said in the introduction, 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, such that if the agent read them, she would lose her belief. Even if the agent doesn't read those sources, their presence can constitute defeaters to her putative knowledge that the dictator died.

In our first variant on Harman's example, 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 literature, and will spend her day reading old novels. 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.

It might be objected here that the agent doesn't know the dictator has died because there are misleading reports around saying the dictator is alive, and she is in no position to rebut them. But this is too high a standard for knowledge. There are millions of people in Australia who know that humans are contributing to global warming on purely testimonial grounds. Many, perhaps even most, of these people would not be able to answer a carefully put together argument that humans are not contributing to global warming, such as an argument that picked various outlying statistics to mislead the reader. And such arguments certainly exist; the conservative parts of the media do as much as they can to play them up. But the mere existence of such arguments doesn't defeat the average person's testimonial knowledge about anthropogenic global warming. Similarly, the mere existence of misleading reports does not defeat our agent's knowledge of the dictator's death, as long as there is no nearby world where she is exposed to the reports. (Thanks here to an anonymous referee.)

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.

### 3 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. Like Russell and Doris, his argument uses a variant of Stanley's Bank Cases.<sup>10</sup> 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  $O$ . In Blome-Tillmann's version of the example, Hannah 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.

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 *reductio* 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  $\neg O$  possibilities on Thursday. She'll be taking them into account in her plans. Someone who is anticipating  $\neg O$  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  $\neg O$  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  $\neg O$  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. 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. In any case, I doubt this is the natural way to read Blome-Tillmann's example. We naturally read Hannah as being rational, and if she is rational she won't have the right kind of confidence to count as knowing  $O$  on Thursday.

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.

## 4 Problematic Conjunctions

Blome-Tillmann offers another argument against IRI, that makes heavy use of the notion of having enough evidence to know something. 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:

(2) John and Paul have exactly the same evidence for  $p$ , but only John has enough evidence to know  $p$ , Paul doesn't.

(Blome-Tillmann, 2009, 328-9)

And this is meant to be a problem, because (2) is intuitively false.

But IRI doesn't entail any such thing. We can see this by looking at a simpler example that illustrates the way 'enough' works.

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, (3) and (4) are true.

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

(4) Ringo can afford the computer.

And (5) is true as well, though there's at least a reading of (6) where it is false.

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

(6) George can afford the computer.

Focus for now on (5). 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 (5) is true. Admittedly there are things close to (5) 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 (5).

Now just like George has enough money to buy the computer, Paul has enough evidence to know that  $p$ . Paul can't know that  $p$ , just like George can't buy the computer, because of his practical situation. But that doesn't mean he doesn't have enough evidence to know it. He clearly does have enough evidence, since he has the same evidence John has, and John knows that  $p$ . So, contra Blome-Tillmann, IRI doesn't entail this problematic conjunction.

In a footnote attached to this, Blome-Tillmann offers a reformulation of 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 (2) 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.' (Blome-Tillmann, 2009, 329n23)

Now 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. So I think this is more of a new objection than a reformulation of the previous point. But might it be a *stronger* objection? Might it be that IRI entails (7), which is false?

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

Actually, it isn't a problem that IRI says that (7) is true. In fact, almost any epistemological theory will imply that conjunctions like that are true. In particular, any epistemological theory that allows for the existence of defeaters which do not supervene on the possession of evidence will imply that conjunctions like (7) are true. For example, anyone who thinks that whether you can know that a barn-like structure is really a barn depends on whether there are non-barns in the neighbourhood that look like the structure you're looking at will think that conjunctions like (7) 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 (7) are not surprising at all.



## 5 Holism and Defeaters

The big lesson of the last few sections is that interests create defeaters. Sometimes an agent can't know  $p$  because adding  $p$  to her stock of beliefs would introduce either incoherence or irrationality. The reason is normally that the agent faces some decision where it is, say, bad to do  $\varphi$ , but good to do  $\varphi$  given  $p$ . In that situation, if she adds  $p$ , she'll either incoherently think that it's bad to do  $\varphi$  although it's good to do it given what is (by her lights) true. Moreover, the IRI theorist says, being incoherent in this way blocks knowledge, so the agent doesn't know  $p$ .

But there are other, more roundabout, ways in which interests can mean that believing  $p$  would entail incoherence. One of these is illustrated by an example 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. That is, we will assume Kate is an expected utility maximiser. 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. Note in particular that street signs are hardly basic epistemic sources. They are the kind of evidence we should be 'conservative' about in the sense of Pryor (2004). We should only use them if we antecedently believe they are correct. So for Kate to believe she's on State, she'd have to believe the street signs she can see are correct. If not, she'd incoherently be relying on a source she doesn't trust, even though it is not a basic source.<sup>11</sup> But if she believes the street signs are correct, she'd believe she was on Main, and that would lead to practical incoherence. So there's no way to coherently add the belief that she's on State Street to her stock of beliefs. So she doesn't know, and can't know, that she's either on State or on Main. This is, in a roundabout way, due to the high stakes Kate faces.

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 only way for the interest-relative theorist to go. What they could, and I think should, say is that Kate can't know she's on State Street because the only grounds for that belief are intimately connected to a proposition that, in virtue of her interests, she needs very large amounts of evidence to believe.

## 6 Non-Consequentialist Cases

None of the replies yet have leaned heavily on the last of the three points from the introduction, 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. But 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 not so clear that orthodox decision theory is the appropriate way to model the agents. That's relevant to cases like this one, which Jessica Brown has argued are problematic for the epistemological theories John Hawthorne and Jason Stanley have recently been defending.<sup>12</sup>

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 I think mistaken, to represent the payoff table associated with 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  $\varepsilon$  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 obligations over and above her duty to maximise expected utility. And that's very implausible, since consequentialism is a fairly implausible theory of medical ethics.<sup>13</sup>

It's not clear exactly what 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  $\delta$ , 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. This is not to say that the surgeon does know that the left kidney is diseased, just that the version of IRI being defended here is neutral on that question.

There is a very general point here. It suffices to derive IRI that we defend principles like the following:

- Whenever maximising expected value is called for, one should maximise expected value conditional on everything one knows.
- Maximising expected value is called for often enough that there exist the kinds of pairs of cases IRI claims exist. That's because in some cases, changing the options facing an agent will make it the case that which live option is best differs from which live option is best given  $p$ , even though the agent antecedently knew  $p$ .

But that doesn't imply that maximising expected value is always called for. Especially in a medical case, it is hard to square an injunction like "Do No Harm!" with a view that one should maximise expected value, since maximising expected value requires treating harms and benefits symmetrically. What would be a problem for the version of IRI defended here was a case with the following four characteristics.

- Maximising expected value is called for in the case.
- Conditional on  $p$ , the action with the highest expected value is  $\varphi$ .
- It would be wrong to do  $\varphi$ .
- The agent knows  $p$ .

It is tempting for the proponent of IRI to resist any attempted counterexample by claiming it is not really a case of knowledge. That might be the right thing to say in Brown's case. But IRI defenders should remember that it is often a good move to deny that the first condition holds. Consequentialism is not an obviously correct theory of decision making in morally fraught situations; purported counterexamples that rely on it can therefore be resisted.

## 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, Jessica. 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. 2010. "Do You Know More When It Matters Less?" *Philosophical Psychology* 23:683–706, doi:10.1080/09515089.2010.514572.

- 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.
- Harman, Gilbert. 1973. *Thought*. Princeton: Princeton University Press.
- Hawthorne, John. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hawthorne, John and Stanley, Jason. 2008. "Knowledge and Action." *Journal of Philosophy* 105:571–90.
- 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.
- Pryor, James. 2000. "The Skeptic and the Dogmatist." *Noûs* 34:517–549, doi:10.1111/0029-4624.00277.
- . 2004. "What's Wrong with Moore's Argument?" *Philosophical Issues* 14:349–378, doi:10.1111/j.1533-6077.2004.00034.x.
- Russell, Gillian and Doris, John M. 2009. "Knowledge by Indifference." *Australasian Journal of Philosophy* 86:429–437, doi:10.1080/00048400802001996.
- 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.

## Notes

<sup>1</sup>It is a tricky exegetical question how many of the three features here must be read into defences of IRI in the literature. My reading is that they do not have to be read in, so it is not overly original of me to defend a version of IRI that does away with all three. But I know many people disagree with that. If they're right, this paper is more original than I think it is, and so I'm rather happy to be wrong. But I'm going to mostly set these exegetical issues aside, and compare different theories without taking a stand on who originally promulgated them.

<sup>2</sup>Such as the Bank Cases in Stanley (2005), or the Train Cases in Fantl and McGrath (2002).

<sup>3</sup>And this is true even though *p* is not a proposition about their interests, or something that is supported by propositions about their interests, and so on.

<sup>4</sup>I will consider, and tentatively support, one principle stronger than IRI in the final section. But the key point is that these general principles are not needed to defend IRI.

<sup>5</sup>This points are expanded upon greatly in (Author Paper 1).

<sup>6</sup>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.

<sup>7</sup>As I believe was the case in the I-35W collapse.

<sup>8</sup>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.

<sup>9</sup>I think the objections I make here are similar in spirit to those Stanley made in a comments thread on [Certain Doubts](http://el-prod.baylor.edu/certain_doubts/?p=616), though the details are new. The thread is at [http://el-prod.baylor.edu/certain\\_doubts/?p=616](http://el-prod.baylor.edu/certain_doubts/?p=616).

<sup>10</sup>In the interests of space, I won't repeat those cases yet again here.

<sup>11</sup>The caveats here about basic sources are to cancel any suggestion that Kate has to antecedently believe that any source is reliable before she uses it. As Pryor (2000) notes, that view is problematic. The view that we only get knowledge from a street sign if we antecedently have reason to trust it is not so implausible.

<sup>12</sup>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.

<sup>13</sup>I'm not saying that consequentialism is wrong as a theory of medical ethics. But if it is right, so many intuitions about medical ethics are going to be mistaken that such intuitions have no evidential force. And Brown's argument relies on intuitions about this case having evidential value. So I think for her argument to work, we have to suppose non-consequentialism about medical ethics.