Belief, Knowledge and Interests

Brian Weatherson

May 8, 2010

Contents

1

Credences and Beliefs, Beliefs and Actions

1 Introduction

In recent philosophical work, there have been interesting debates over these questions.

1. What is the relationship between the subjects of traditional epistemology (e.g., rational belief, knowledge, etc.) and rational action?
2. What is the relationship between the subjects of traditional philosophy of mind (e.g., beliefs, desires, etc.) and the subjects of modern Bayesian epistemology such as credences, utility functions and preferences?

On the first question, John Hawthorne, Jason Stanley and Jeremy Fantl and Matthew McGrath have argued that a necessary condition on something being a piece knowledge is that it can ground rational action (Hawthorne, 2004; Stanley, 2005a; Hawthorne and Stanley, 2008; Fantl and McGrath, 2009). Hawthorne and Stanley go further, and say that only knowledge can provide a certain kind of grounding for rational action. This kind of constraint implies, given some plausible assumptions, that knowledge is interest-relative.

On the second question, various philosophers have argued that beliefs are just special kinds of credences. In particular, to believe that p is simply to have a credence in p above some threshold. This kind of view is defended by Richard Foley (1993), David Hunter (1996), David Christensen (2005) and Scott Sturgeon (2008). I'm going to argue that this position isn't quite right. Just as the epistemologists mentioned in the previous paragraph argued that we should *evaluate* binary states like belief using interest-relative standards, I'm going to argue that we should *identify* beliefs in an interest-relative way.

It should be clear why the two questions are related. Bayesian theory suggests a strong connection between rational credences and utility functions, on the one hand, and rational action, on the other. If we think that answer is correct, then the 'product' of the answers we give to questions 1 and 2 must match up with it. Now the answer the Bayesians give, that rational agents maximise expected utility, is controversial, and may not be appropriate in all circumstances. But it seems very plausible that in *many* circumstances there is *some* definition of utility such that rational action coincides with maximising expected utility. If we identify beliefs with certain kinds of credences, then say something about what role rational beliefs can play in

the rationalisation of action, we might have ended up committing ourselves to the rationality of actions that don't maximise expected utility.

Here's an example of how that could happen. Let's say that we identify belief with credence greater than 0.95, and rational belief with rational credence greater than 0.95. And let's say we adopt the principle that if S rationally believes p, and rationally believes that if p is true, then ϕ will have the best outcome of any available action, then it is rationally permissible to ϕ. Then we'll end up saying that agents need not maximise expected utility in cases like the following.

Bad Roulette: S is facing a known to be fair roulette wheel that's about to be spun. She's holding a chip that will be worth \$100 if the ball lands on 17. Carrying the chip has some nominal costs. (It's better to not carry useless things, so anything has a small carrying cost.) Someone offers to take the chip off her hands; i.e, to carry the chip in exchange for getting the \$100 if the ball lands on 17. Let p be the proposition that the ball won't land on 17. S is rational, so her credence in it is $^{36}/_{37}$, which is greater than 0.95. By hypothesis, that means she believes p. And she knows that if p, she's best off taking the stranger's offer to carry the chip. So, by hypothesis, she should take the offer. But that's crazy. Or, at least, it doesn't maximise expected utility, which for present purposes we're treating as the same thing.

So one of these two answers must be wrong. As it turns out, I'm going to argue that both answers are wrong.

What I hope is clear from this example is that certain combinations of answers to the two questions are ruled out. In this case neither element of the combination is clearly absurd. Indeed, the proposed answer to question 1 is taken straight from Fantl and McGrath (2009), and the proposed answer to question 2 is in the spirit of the views of Foley, Hunter, Christensen and Sturgeon. But I think, and I hope this isn't an unfair reading, that none of these authors offers comprehensive answers to questions 1 and 2. If I'm right that an answer to either of these questions has to say something plausible about cases like **Bad Roulette**, then this is an important oversight. And the best way to address it is to try to answer both questions at once, as I'll do in this work.

Bad Roulette should be a familiar *kind* of case to those who have been following the literature on contextualism and interest-relative invariantism over the past decade or so. It has a very similar structure to the Bank Cases (Stanley, 2005a; DeRose, 2009), Airport Cases (Cohen, 1999), Train Cases (Fantl and McGrath, 2009) and the like that have dominated that literature. What's common to each of these cases is that there is a proposition and a choice such that (a) the agent's evidence strongly supports the proposition, and (b) if the proposition is true, the best results will come from taking one option, but (c) the agent maximises expected utility by taking some other option. For instance, in the Bank cases, the bank will very likely be open the

next day, and if it is, the utility maximising option will be to return the next day.[1] The upshot is that we can't identify both identify rational belief with rational high credence, and say that rational belief can directly support rational action.

But I want to stress two respects in which the role that **Bad Roulette** plays in my argument is unlike the role that those cases have played in earlier arguments for contextualism.

First, what makes Bad Roulette distinctive is not that it is involves 'High Stakes' in any particular way. \$100 is not nothing, but it isn't an amount that should trigger any super-high epistemic standards on its own. What matters is the *odds* at which the agent is asked to bet on p. If S accepts the offer, she gains something worth maybe a penny is p is true, and loses \$100 if p is false. In effect she's asked to bet on p at odds of 10,000:1. That's a bad bet even at low odds.

Second, I don't claim that it is obvious, either intuitively or reflectively, what to say about Bad Roulette. I'm not arguing from facts about that case, or intuitions about it. My starting point instead is that expected utility maximisation, i.e., orthodox decision theory, is at least some of the time the right theory of rational action for some natural measure of utility. We can stipulate that S's cares, reasons and values are such that she should maximise some gently declining function of her material assets. So giving away her chip will definitely not maximise expected utility. So she shouldn't do it.

Let's go through a pair of cases, which I'll call the Map Cases, that make both of these points explicit.

High Cost Map: Brian 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 Brian 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

[1] Or at least that's how the Bank Cases are usually told. Personally, I'm not convinced it is true. I've rarely seen a bank line so long that it's worth driving back to the bank the next day. I suspect that most people, including most people reading and writing these cases, overestimate the disutility of queueing, and underestimate the disutility of driving. The Train Cases are more clearly of this form.

> 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 Brian 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 Brian'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, I'm going to argue in subsequent chapters, 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. And I don't think the case is particularly intuitive. I'm happy to *accept* a theory that says Brian knows in one case but not the other, but I don't think this is at all obvious. This shouldn't be too surprising. When it comes to reasoning about probability, we all tend to be susceptible to various errors. When we're going by snap judgments, or intuitions, we're even less reliable. When we're thinking about 'tail risk', or the importance of obscure and hard to price events, we're less reliable still. In short, this is not an area where we should think native intelligence is a particularly trustworthy guide to the truth. That's not to say that we should let our theories have implications that are obviously false. Indeed, we shouldn't let our theories have implications that are non-obviously false. It's just that in this particular field we have lots of evidence that fewer things are indeed obvious than in other fields.

2 Assumptions

Throughout this work, I'm going to make two prominent assumptions.

The first of these I've already flagged. It's the assumption that orthodox expected utility theory is, at least some of the time, a good theory of rational action. So, at least some of the time, the fact that an agent does not maximise expected utility implies that they are not acting rationally.

This assumption needs to be clarified and qualified in a couple of respects. I mean to be endorsing *causal*, not *evidential* decision theory here, though I don't think anything here will turn on that. More substantively, I don't think that we should *always* maximise expected utility. Arguably there are some moral constraints that are not reducible to constraints on the structure of the utility function.[2] And even when we should maximise some function, it won't be clear just which function that

[2]There's a big debate on this point that I'm not going to get into here. See Portmore (2009) for more details and references.

is. I don't assume that 'utility' always measures welfare, let alone some component of welfare like preference-satisfaction. But I think in some cases, like the three cases in the previous section, it is plausible that the agent should simply be trying to maximise their own expected welfare. That's not because of any broad theory about the nature of good action; it's simply that nothing else is relevant to their decisions for the next few minutes. As we'll see quite a bit in what follows, in cases where more than the agent's short-term welfare turns on a decision, describing and evaluating that decision can get more complicated.

The term 'expected' in the reference to expected utility also calls for some unpacking. I assume that, at least some of the time, there is such a thing as the probability of various hypotheses given the agent's evidence. I don't assume this is always the case.[3] Nor do I assume this is always precise.[4] But I do assume that it is well-enough defined in enough cases that we can sensibly talk about the expected utility of various actions for an agent.

The second assumption is that some kind of functionalism about the mind is broadly true. In particular, I'm going to assume that the theory sketched in part one of Lewis (1994), and filled out in much more detail in Braddon-Mitchell and Jackson (2007) is correct.Lewis (1994) says that he's not sure whether his view falls into the extension of 'functionalist' as that term is often used by philosophers. Whether that's true or not, I'm using 'functionalist' in such a way that Lewis's theory is a functionalist theory. Note that both Lewis and Braddon-Mitchell and Jackson do not take functionalism to be incompatible with an identity theory, and I won't either. Indeed, I'm going to make use of the idea that mental states are just brain states, and the role of functionalism is to tell us which functional role a brain state must fill to be a mental state, at a key point of the argument. I'm making these assumptions because I think they're true, and I think they provide a crucial insight into how we should answer the two questions I started with. It might be worried that readers less enamoured of functionalism will not find a work on epistemology for functionalists particularly interesting. I think such a worry is misplaced. Some people might find it interesting in the way that psychologists find morality for psychopaths interesting. More seriously, some readers might agree with me that if functionalism is true, the right epistemology has a certain structure, and take that to be evidence against functionalism. More hopefully, the fact that functionalist assumptions help resolve some tensions in modern epistemology might increase the credibility of functionalism in some people's eyes.

It's important to the kind of functionalism that I'm adopting that mental states have three kinds of proprietary functional roles. Mental states are associated with (a) input conditions, (b) output conditions, and (c) internal connections. So for a state to be a belief that p, it should be (a) produced by evidence that indicates p, (b) support the formation of other beliefs whose contents are supported by p, and desires that are rational in a world where p is true, and (c) lead to action that maximises the agent's

[3] See Weatherson (2007) for some discussion of why we shouldn't think things like epistemic probabilities are well defined in situations where the agent has very little evidence.

[4] But see Weatherson (2002) for some discussion of why we can act as if probabilities are precise in at least some cases where they are not.

interests conditional on p being true. This is something of an idealisation; not all beliefs behave in just this way. But if a state rarely has one of these features among the class of creatures that includes the agent, then the functionalist will deny that it is really a belief.

In an earlier version of this work (Weatherson, 2005a), I focussed on the second and third of these conditions. That led some writers to say that I was really talking about 'acceptance', not about 'belief'.[5] I think that's because when people talk about acceptances, they mean mental states that satisfy clauses (b) or (c), or perhaps both, in the functionalist account of belief. Consider, for example, the case raised by L. Jonathan Cohen (1989, 369) of the lawyer who doesn't believe his client is guilty, but organises the preparation of the defence around that assumption. The lawyer's state isn't a belief, presumably, because it doesn't have the right relationship to the evidence.[6]

3 Functions of Credences and Beliefs

Frank **?** provides a clear statement of one of the functional roles of credences; their connection to action. Of course, Ramsey did not take himself to be providing one-third of the functional theory of credence. He took himself to be providing a behaviourist/operationalist reduction of credences to dispositions. But we do not have to share Ramsey's metaphysics to use his key ideas. I'm going to focus on two ideas of Ramsey's here. First, the idea that it's distinctively *betting* dispositions that are crucial to the account of credence. And second, the idea that all sorts of actions in everyday life constitute bets.

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

The functionalist goes along with much, but not quite all, of this picture. The betting preferences are an important part of the functional role of a credence; indeed, they just are the output conditions. But there are two other parts to a func-

[5] This seems to be what is suggested by (Fantl and McGrath, 2009, 149-151), for instance.

[6] I'm actually a little suspicious that the case as described really works as a case of acceptance without belief. If the lawyer really acted as if the client is innocent, there are probably a few things he would do differently to how he would actually act. He might prepare more lines of appeal, because it's worth doing that for an innocent client. If it's a serious charge, he might enlist the support of academic and advocacy groups that help the wrongly accused. He might be more willing to have his child visit him at work while the client is there, etc. In practice it is rather rare for us to act exactly as if p is true, where p is relevant to action, and we have little reason to believe p is true. It's possible, which is enough for the theoretical distinction between acceptance and belief to be useful, but I think its frequency is sometimes overstated.

[7] See **?** for the most developed account in recent times.

tional role: an input condition and a set of internal connections. So the functionalist thinks that the betting dispositions are not quite sufficient for having credences. A pre-programmed automaton might have dispositions to accept (or at least move as if accepting) various bets, but this will not be enough for credences. More relevant to our purposes, a ...

4 The Interest-Relativity of Belief

In the previous section we argued that for rational agents, conditionalising on something they believe doesn't change the correct answer to any inquiry. And what's true for rational agents is, as usual in a functionalist story, part of the functional role that individuates mental states. So what it is for an agent to believe that p is, in part, for the agent to not change their mind when they conditionalise on p. As we'll see presently, this looks much too strong. For just about any proposition, there's *some* question whose answer looks different when you conditionalise on that proposition. So it looks like we believe very little. That, I think, would be a mistaken conclusion. The better way forward is to restrict the quantifier 'any question'. What's really true is that if a rational agent believes p, then conditionalising on p doesn't change the answer to any *relevant* question, where what is relevant depends on their interests.

This, I propose, is a key way in which beliefs are interest-relative. What we believe is in part a function of what we care about. In Weatherson (2005a) I argued that this explains *all* the interest-relativity we find in epistemology. For reasons I'll get to later in the chapter, I no longer believe that. But I do believe it explains a lot of the interest-relativity. I'm somewhat confident that it explains all the interest-relativity that has been hitherto uncovererd in the literature, for example. But before we get to that conclusion, let's step back to see why interests need to come in.

Beliefs structure inquiry and planning. If you believe p, then p is a fixed point, a given, around which you plan. There is perhaps a more casual sense of 'believe' in which this isn't true. If Smith says, "I believe I turned the gas off, but I better go back and check," then Smith does *not* believe, in the sense we're using here, that she turned the gas off. If she did believe that, there wouldn't be any need to go back and check, since she would have settled to her own satisfaction the question of whether the gas was turned off. The sense of 'believe' I'm using here, where to believe something is to take it to be settled, is I think the common philosophical usage in epistemology and philosophy of mind. When we talk about the conditions necessary for a belief to be justified, we don't mean to talk about what would justify an agent being as confident in a proposition as Smith is that she turned the gas off. We mean to be aiming for something stronger. Or consider what we mean when we say that an agent who wants ϕ, and believes that doing ψ is the best way to get ϕ, should indeed do ψ. We don't mean that if the agent is kinda confident that doing ψ is the best way to get ϕ, she should do ψ. Rather, we mean that if the agent has settled to their own satisfaction that doing ψ is the best way to get ϕ, then ψ is what they should do. On this point, I'm going to follow standard philosophical usage throughout this work.[8]

[8]There is some interesting discussion in (Fantl and McGrath, 2009, ??) on the relationship between the casual and the philosophical usages.

What we take as given in inquiry varies as the subject of inquiry varies. The easiest way to see this is by thinking of the examples John Hawthorne presents at the start of Hawthorne *Knowledge and Lotteries*. Imagine first that I'm inquiring about what I'll be teaching next Fall, where I know this is information that is accessible to me.[9] In such an inquiry, I'll simply assume that I won't be dead by next Fall, or have been fired, or that my school will have folded for financial reasons. Rather, I'll just focus on what I know about the teaching schedule. Now imagine that I'm deciding whether to buy either life insurance or unemployment insurance, so I'm trying to figure out the likelihood that such a policy will pay out. In that inquiry, I clearly won't assume that I'll be alive and teaching in a year's time. Indeed, the point of the inquiry is to figure out whether just those propositions, the ones I had previously taken as given, are actually true. Note the claim here is descriptive, not normative. I'm not saying yet that I should take different things as given in different inquiries. I happen to believe that, but I'm not arguing yet. All I'm relying on so far is the observation that I do make different presuppositions in different inquiries, and I assume that you do too.

The last two paragraphs suggest an argument that we believe very little. If a belief should be presupposed in any inquiry, and very little is presupposed in all inquiries, then we believe very little. We can put the same point in the terms used in the previous section. For almost any proposition p, there is some inquiry where conditionalising on p changes the answer. Most obviously, there is the inquiry into the probability of p. Since the conditional probability of p given p is 1, conditionalising on p will change the answer to the question *What is the probability of p?* unless the answer to that question was already 1. So only propositions whose probability is 1 are believed. But that's not even a sufficient condition for belief. We might also be interested in what the conditional probability is of p given $\neg p \vee q$, where q is some absurdity. Even if we think the probability of p is 1, we might think the conditional probability of p given $\neg p \vee q$ is less than 1. (That is, if we're either wrong about p, or wrong about $\neg q$, we might think it at least possible that we're wrong about $\neg q$.) But the conditional probability of p given $p \wedge (\neg p \vee q)$ is, of course, 1. So if conditionalising on p does not change the answer to *any* question, it must be that for any q, the probability of p given $\neg p \vee q$ is 1, for just about any q.[10] And very few propositions, outside perhaps of basic arithmetical truths, satisfy that condition.

A natural thought here would be that the model of the previous section is wrong. After all, we believe a lot, but we've just seen an argument that given that model, we believe very little. So, we might well conclusion, the model must be wrong. That's not what I'm going to conclude, however. I'm going to conclude that the model has been misapplied, and that the correct application takes us to the truth.

Let's focus again on the platitudes about belief we developed. Beliefs structure inquiry. Different inquiries have different structures. A natural inference to draw

[9]The inquiry might just involve racking my brains for the relevant memory. When I talk about inquiries, I don't mean anything nearly as formal as a judicial inquest, or even something that need take more than a few seconds.

[10]There are technical issues that arise here if q is an impossibility. I mean to simply side-step those issues, which is why the weaselly phrase 'just about' is there.

from these platitudes is that we believe different things when engaged in different inquiries. And we can put this into the formal model. We said that to believe p it must be that conditionalising on p does not change the answer to any question. We should have qualified that by saying that the question must be one that's relevant to your current inquiries. You can believe p, even though you wouldn't take p as given in a different inquiry. All that qualification shows is that if you were engaged in a different inquiry, you wouldn't believe p. Since you're not engaged in that inquiry, it's possible you do believe p.

It's time to say a little bit more about what it is to be engaged in an inquiry. I'm not going to offer a short snappy answer to this question, but I will fill in some key details. We'll start with practical inquiries, in particular inquiries about whether to do ϕ or ψ, where these are actions. An action ϕ is a live option for the agent if it is really possible for the agent to perform ϕ. An action ϕ is a salient option if it is an option the agent takes seriously in deliberation. Most of the time gambling large sums of money on internet gambling sites over my phone is a live option, but not a salient option. I know this option is suboptimal, and I don't have to recompute every time whether I should do it. Whenever I'm making a decision, I don't have to add in to the list of choices *bet thousands of dollars on internet gambling sites*, and then rerule that out every time. I just don't consider that option, and for what it's worth, I'm right do so. If I have a propensity to daydream, then it might be that becoming the centrefielder for the Boston Red Sox is a salient option to me, but it certainly isn't a live option. Whenever ϕ and ψ are live and salient options for the agent, the question of which of them to do is part of her inquiries. So if she believes p, and ϕ and ψ are live, salient options, her preference ranking over them must equal her preference ranking over them given p.

An agent need not only make decisions about what to do; she might also decide what to do conditional on some hypothesis or other. We need to factor this into our definition of an inquiry. Say a proposition is *relevant* if the agent is disposed to take seriously the question of whether it is true (whether or not she is currently considering that question) and conditionalising on that proposition or its negation changes some of the agents *unconditional* preferences over live, salient options.[11] The first clause is designed to rule out wild hypotheses that the agent does not take at all seriously. If q is not such a proposition, if the agent is disposed to take it seriously, then it is relevant if there are live, salient ϕ and ψ such that $\phi \geq_q \psi \leftrightarrow \phi \geq \psi$ is false. Say a proposition is *salient* if the agent is currently considering whether it is true. Finally, say a proposition is *active* relative to p iff it is a (possibly degenerate) conjunction of propositions such that each conjunct is either relevant or salient, and such that the conjunction is consistent with p. (By a degenerate conjunction I mean a conjunction with just one conjunct. The consistency requirement is there because it might be hard in some cases to make sense of preferences given inconsistencies.) Then if q is active, and phi and psi are live and salient, the question of whether ϕ

[11] Conditionalising on the proposition *There are space aliens about to come down and kill all the people writing epistemology books* will make me prefer to stop writing this paper, and perhaps grab some old metaphysics papers I could be working on. So that proposition satisfies the second clause of the definition of relevance. But it clearly doesn't satisfy the first clause, so it's not relevant.

or psi is better given q is part of the agent's inquiry. So the agent only believes p if $\phi \geq_q \psi \leftrightarrow \phi \geq_{p \wedge q} \psi$.

Not all inquiries are practical inquiries. Most of the inquiries that this book concerns are not, for example. These theoretical inquiries also affect what is relevant to the agent. One of the shortcomings of previous work on interest-relativity in epistemology is that the various author's haven't spent as much time as they should on *theoretical* interests of inquirers. Let's say a bit about those.

Some inquiries are inquiries into the truth of some proposition. When the agent is making such an inquiry, she doesn't take the answer as given. So when the question *Is it true that p?* is a live question for the agent, she doesn't believe that p until that question is answered to her satisfaction. Our model captures this in a somewhat roundabout way. Since the question *p?* is live, but the question *Given p, p?* is not live (assuming minimal rationality), the agent has different attitudes to a question, namely *p?*, depending on whether or not we conditionalise on p. So she doesn't believe p.

It's more interesting to think about inquiries into the probability of some proposition. We can ask all sorts of questions about the probability of various propositions. In some cases we might be interested in these questions for practical reasons, such as when we are deciding whether to buy some kind of insurance. But in other cases we care about them simply out of curiosity, such as when we wonder about the probability of human life persisting for another 10,000 years. In these cases, a question is relevant to the agent simply if she makes it relevant by thinking about it.

But we should be careful about just what kind of probabilistic questions agents ask. It's actually somewhat uncommon to be concerned with the *exact* value of some probability. It's much more usual to be concerned with whether a probability falls in some interval, or whether one of two propositions is more probable than another. We might ask the question about intervals directly, as in *Is the probability of p greater than 0.95?*. Or we might ask it somewhat indirectly, as when we ask what the probability of p is, but we really only care about the first two significant figures. In either of these cases, a bit surprisingly, it is possible to believe that p even though the probability of p is less than 1. That's because whether the probability of p falls into a relevant interval doesn't change when we conditionalise on p. That's true in the first case (*Is the probability of p greater than 0.95?*) as long as p's unconditional probability is greater than 0.95, and in the second case (*What's the probability of p to two significant figures?*) as long as the probability of p is greater than 0.995. So even raising probabilistic questions about p isn't, strictly speaking, incompatible with believing p when p isn't absolutely certain.[12]

Let's sum all this up. A key part of the functional role of belief is that beliefs structure inquiry. But how an agent structures inquiry depends on what they're inquiring into. I draw the obvious conclusion that what an agent believes depends on what they're inquiring into. We might be able to turn all this into a reductive analysis.

[12]There's a surprising asymmetry around here. If we ask what the probability of $\neg p$ is to some number of significant figures, the answer to that question will be different depending on whether we conditionalise on p unless the unconditional probability of p is 1. In some ways, that's the more natural case. It's a little unnatural to take p as given in an inquiry into the probability of p, and it's a little surprising that our model allows this.

The 'structuring' role of belief captures a lot of the internal role of belief, and belief's role in generating action. It doesn't cover the 'input' conditions for belief. But the input conditions for belief, that beliefs are based in evidence, are the same as the input conditions for credences. So the following account of belief arguably captures all the functional roles of belief.

> S believes that p iff for any inquiry S is engaged in, S's credence in p is high enough that conditionalising on p will not change the answer she gives to that inquiry.

5 Two Caveats

The theory sketched in the previous section seems to me right in the vast majority of cases. It fits in well with a broadly functionalist view of the mind, and as we'll see it handles some otherwise difficult cases with aplomb. But it needs to be supplemented and clarified a little to handle some difficult cases. In this section I'm going to supplement the theory a little to handle what I call 'impractical propositions', and clarify a little the notion of conditional answers.

Jones has a false geographic belief: he believes that Los Angeles is west of Reno, Nevada.[13] This isn't because he's ever thought about the question. Rather, he's just disposed to say "Of course" if someone asks, "Is Los Angeles west of Reno?" That disposition has never been triggered, because no one's ever bothered to ask him this. Call the proposition that Los Angeles is west of Reno p.

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

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

Here's a way out of this problem. We supplement the theory of the previous section with these principles.

- A proposition p is *eligible for belief* iff it satisfies the theory of the previous section. That is, p is eligible for belief iff conditionalising on p does not change the answer to any inquiry the agent is engaged in.

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

- For any proposition p, and any proposition q that is relevant or salient, among the actions that are (by stipulation!) open and salient with respect to p are *believing that p*, *believing that q*, *not believing that p* and *not believing that q*
- For any proposition, the subject prefers believing it to not believing it iff (a) it is eligible for belief and (b) the agents degree of belief in the proposition is greater than 1/2.
- The previous stipulation holds both unconditionally and conditional on p, for any p.

This all looks moderately complicated, but I'll explain how it works in some detail as we go along. One simple consequence is that an agent only believes that p iff their degree of belief in p is greater than 1/2. Since the schoolboy's degree of belief that Los Angeles is west of Reno is not greater than 1/2, in fact it is considerably less, he doesn't believe $\neg p$. On the other hand, since his degree of belief in p is considerably greater than 1/2, he prefers to believe it than disbelieve it, so he believes it.

There are many possible objections to this position, which I'll address sequentially.

Objection: Even if one has a high degree of belief in p, one might prefer to not believe p because one thinks that belief in p is bad for some other reason. Perhaps, if p is a proposition about one's own brilliance, it might be immodest to believe that p.
Reply: Any of these kinds of considerations should be put into the credences. If it is immodest to believe that you are a great philosopher, it is equally immodest to believe to a high degree that you are a great philosopher.

Objection: Belief that p is not an action in the ordinary sense of the term, so it's wrong to say that one of the *actions* available to the agent is believing that p.
Reply: True, which is why this is described as a supplement to the original theory, rather than just cashing out its consequences.

Objection: It is impossible to choose to believe or not believe something, so we shouldn't be applying these kinds of criteria.
Reply: I'm not as convinced of the impossibility of belief by choice as others are, for reasons set out in **?** but I won't push that for present purposes. Let's grant that beliefs are always involuntary. So these 'actions' aren't open actions in any interesting sense, and the theory of the previous section was really incomplete. As I said, this is a supplement to that theory.

It might be thought that if choices about beliefs are impossible, it's impossible for them to be 'rational' or 'irrational', so the functionalist's characteristic move of individuating states in everyone by the way those states behave in the ideal agent won't be available. But of course if some kind of state is unavoidable, the ideal agent doesn't avoid it. So we can still harness features of the ideal agent to functionalist ends.

Objection: This just looks like a roundabout way of stipulating that to believe that p, your degree of belief in p has to be greater than 1/2. Why not just add that as

an extra clause than going through these little understood detours about preferences about beliefs?

Reply: There are three reasons for doing things this way rather than adding such a clause.

First, it's nice to have a systematic theory rather than a theory with an ad hoc clause like that.

Second, the effect of this constraint is much more than to restrict belief to propositions whose credence is greater than 1/2. Consider a case where p and q and their conjunction are all salient, p and q are probabilistically independent, and the agent's credence in each is 0.7. Assume also that p, q and $p \wedge q$ are completely irrelevant to any practical deliberation the agent must make. Then the criteria above imply that the agent does not believe that p or that q. The reason is that the agent's credence in $p \wedge q$ is 0.49, so she prefers to not believe $p \wedge q$. But conditional on p, her credence in $p \wedge q$ is 0.7, so she prefers to believe it. So conditionalising on p does change her preferences with respect to believing $p \wedge q$, so she doesn't believe p. So the effect of these stipulations rules out much more than just belief in propositions whose credence is below 1/2.

This suggests the third, and most important point. If we just stipulated that belief requires credence greater than 1/2, we won't be able to support some plausible closure principles for rational belief. We'll return to those principles in (much) greater length when we discuss the 'threshold view' of belief.

The other caveat to the theory is that it only works if we make a slightly non-standard assumption about conditional values in cases where an agent's action can harm others. I think the assumption is plausible; indeed, I think it is true. But without the assumption, the theory will have some bad results in the kinds of cases discussed by Jessica ?. [14]

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 ?.

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.

[14] The following point is I think of some intrinsic interest, and will be crucial to what we say about some examples later on. But it does get a little deep in the decision-theoretic weeds, and some readers may prefer to skip to the next section.

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

	p	$\neg p$
Do ϕ	-10	0
Do ψ	0	-10
Find out which is harmful	-1	-1

Given that $Pr(p) = 1/2$, it's easy to compute that the expected value of doing either ϕ or ψ is -5, while the expected value of finding out which is harmful is -1, so the agent should find out which thing is to be done before acting. So far most consequentialists would agree, and so probably would most non-consequentialists for most ways of fleshing out the abstract example I've described.[15]

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

In the abstract statement of the situation above, I said that one of the options would be *harmful*, but I didn't say who it would be harmful to. I think this matters. I think what I called the orthodox treatment of the situation is correct. 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

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

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 ϕ 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.

The theory we end up with from all these reflections is somewhat hard to classify. On the one hand, the examples we've been looking at are ones where the values of various actions are all based on the value of various possible consequences. From that perspective the theory feels somewhat consequentialist. On the other hand, we've divorced good action from good consequences fairly sharply. We've got a situation where an agent doesn't know what action will have the best consequences, but does know what action has the highest value. I'll leave the question of whether this makes the theory consequentialist or non-consequentialist, or indeed the question of whether anything turns on this classification, to others. What's crucial to us is that we're assuming this theory is true. We'll have cause to return to it when discussing the effect of conditionalisation on cases where an agent may harm others she has various duties towards. As you might expect, the distinctive result we'll draw will be that conditionalisation has less effect than orthodox theory would suggest in these cases, since the agent was really making a decision under certainty.

6 Summing Up

2

Defending the Interest-Relative Theory of Belief

1 Sturgeon's Challenge

2 Thresholds, Actions and Significance

Both of the objections I'm going to make to the Threshold View are found in a short passage by Robert Stalnaker (1984, 91). Neither objection seems to have been particularly convincing to the wider world, since the Threshold View retains, and even gains, adherents. So I'll spend some time saying why each of them are good objections.

The Threshold View says that S believes that p iff S's credence in p is greater than some salient number r, where r is made salient either by the context of belief ascription, or the context that S is in. As Stalnaker emphasises, any number r is bound to seem arbitrary. Unless these numbers are made salient by the environment, there is no special difference between believing p to degree 0.9786 and believing it to degree 0.9875. But if r is 0.98755, this will be *the difference* between believing p and not believing it. Intuitively, that difference *is* an important difference, so it can't simply be the difference between believing p to degree 0.9786 and believing it to degree 0.9875.

The usual response to Stalnaker, as found in Foley (1993, Ch. 4) and Hunter (1996) is to say that the boundary is vague. But when we consider in more detail how vagueness is supposed to help here, this response doesn't seem successful. That's because on many different theories of vagueness, whatever is true on any way of making the vague term precise is still true. Or, at the very least, anything that isn't expressly about the vagueness of the term and is true on all precisifications is also true. And we have good reasons to think that some such theory of vagueness must be true.

If epistemicism about vagueness, as defended by for example Timothy Williamson (1994), is correct, then Foley and Hunter's position seems untenable. If epistemicism is correct, then every vague term picks out a boundary that is metaphysically precise, but unknowable. But the problem with the arbitrariness of r is not solved by noting that we can't know where r is. The same kind of conclusion will follow on any theory of vagueness where boundaries are metaphysically sharp, such as the contextualism defended by Delia Graff Fara (2000).

If a traditional degree of truth theory, as defended by for example Kenton Machina (1976) is correct, this objection won't go through. After all, it isn't true on such a theory that anything that's true on any precisification is simply true. For example, the claim that S doesn't both believe and not believe p isn't true on the degree of truth theory if it's true to degree 0.5 that S believes that p. But this leads to two new problems. One technical problem is that it isn't clear this is still the Threshold View, since it isn't true that there's a threshold.[1] But I won't stress this, since the view is obviously in the spirit of the Threshold View. The bigger problem, as pointed out by Timothy Williamson (1994, Ch. 4) is that it is very implausible that these negated conjunctions are not true. Nothing in the nature of vagueness, argues Williamson, gives us reason to deny that contradictions are always false. Modern degree of truth theories, such as the theory I defend in Weatherson (2005b), agree with Williamson that all contradictions should be false. But these theories also agree with epistemicism that what's true on all precisifications is really true.

A little surprisingly, the theory of vagueness that looks most promising for the Threshold View is supervaluationism. This is surprising because you might think that the core premise of my objection, that what is true on any precisification is simply true, is the defining characteristic of supervaluationism. But matters turn out to be a little more delicate than that. After all, supervaluationists have learned to say that some things that are true on all precisifications aren't necessarily true.[2] For instance, it's plausibly true on all precisifications that 'bald' is precise. It doesn't follow that it's really true that 'bald' is precise. So meta-linguistic claims that are true on all precisifications need not be true.

If we were being clumsy with the objection to the Threshold View, it might seem like it turns on a metalinguistic claim. The clumsy way of stating the objection would be to say that it's implausible that the difference between things that satisfy 'believes' and things that don't is arbitrary. If that were the objection, supervaluationism could offer a way out. That's because it's possible to imagine a combination of the Threshold View and supervaluationism which says that although it's true on every precisification that 'believes' denotes an arbitrary boundary, it isn't actually true that 'believes' denotes an arbitrary boundary.

The way around this problem is to note that we can state the objection without *mentioning* any of the key terms. The core intuition is that the difference between having and not having a belief is, in general, a more psychologically significant distinction than any particular difference between nearby credences. That intuition uses terms like 'belief' and 'credence', it doesn't mention them. So if it's false, according to the Threshold View, on every precisification, then it is simply false on the Threshold View. And that's an unacceptable result.

We don't need to simply appeal to the popularity of these theories of vagueness to argue against the Threshold View here. It's independently plausible that the correct

[1] That is, the existentially quantified claim $\exists r \forall S \forall p (S$ believes that p iff S's credence in p is greater than $r)$ will have a very low degree of truth. That is because on these degree of truth theories, the truth value of an existentially quantified claim is the maximum of the truth value of the various instances of that claim, and no instance of this claim will be more than minimally true.

[2] I think the best discussion of this point is still Lewis (1993).

theory of vagueness should respect the principle that whatever's true on all precisifications is true. A little noticed reason for accepting this claim is that it plays a key role in certain parts of social science. I can illustrate this with an example from one of the most discussed academic books of the last hundred years, John Maynard Keynes's *The General Theory of Employment, Interest and Money* Keynes (1936).

The following five claims are all defended in the *General Theory*, the first two on page 61, and the latter 3 on pages 225-6.

(1) All goods are (definitely) investment goods or consumption goods.
(2) For some goods it is vague whether they are an investment or consumption good.
(3) The yield of an investment, q, is vague.
(4) The carrying cost of an investment, c, is vague.
(5) The net yield of an investment, $q - c$, can be precisely determined.

Since Keynes endorsed these claims, he also endorsed, at least tacitly, their consistency. The *General Theory* has been one of the most closely studied, and carefully critiqued, books of our time. And, to the best of my knowledge, the assumptions that these claims were consistent has not been criticised.[3]

This is a problem for theories of vagueness that do not allow that these claims are consistent. On Machina's theory, (1) and (2) are not consistent, since a disjunction like (1) can only be definitely true if one disjunct is definitely true, and (2) denies that either disjunct in (1) is definitely true. So some philosophers (though apparently no economists) disagree with Keynes on the consistency of (1) through (5). Since so many economists have accepted the consistency of these claims, we can safely say that it is intuitive that they are consistent. So a theory of vagueness that takes learned intuitions to be evidentially significant should agree that the claims are consistent. Moreover, the fact that things like net yields can be precise even though their components (gross yields and carrying costs) are vague plays a crucial role in Keynes's theory, and hence in 20^{th} century economics. So a theory of vagueness that takes embedding in a flourishing science to be evidentially significant should agree that these claims are consistent. I'm inclined to think both that intuitions about vagueness are useful evidence, and that embedding in a flourishing science is useful evidence, so I think considering things like consumption goods and net yields gives us very good evidence against theories that deny the consistency of (1) through (5).

And note that it's not just that Keynes says that (1) through (5) are consistent. Both on page 61 and again on pages 225-6, he says that the reason he takes them to be consistent is that he thinks we should take to be true whatever turns out to be true on any arbitrary way of precisifying a vague boundary. So these passages, and their acceptance by a wide swathe of economists with widely differing views on the broader Keynesian program, provide strong evidence that whatever is true on any way of precisifying a vague boundary is simply true.

[3]Not even in books like Coates (1996) that are explicitly about the role of vagueness in the *General Theory*.

All of this is to say that if there's a problem with thresholds being arbitrary, making the thresholds vague seems to be of no help at all. Indeed, it's arguable that the vagueness of the threshold makes it even more arbitrary in some respects. The lesson of these examples from economics is that we don't go wrong by treating vague distinctions as arbitrary distinctions. So perhaps the defender of the Threshold View should deny that there's a problem with the distinction between belief and non-belief being arbitrary.

We can see that this position is unsustainable by thinking again about the role of belief in leading to action. Let's recall just what the Threshold View has to say about the decisions that will be reached in a case like *Bad Roulette*. Recall that *S* receives an offer that has a clearly negative expected utility, but which has a very high probability of having a small positive return. Since *S*'s credence that the ticket will lose is high, according to the Threshold View, she believes her ticket will lose. So she believes that taking the offer will produce the best outcome. And she only cares about getting the best outcome, at least among the available choices. But she is not at all motivated to take the offer. This is, at the very least, rather strange.

And I think the strangeness is grounded in the arbitrariness of the Threshold. The difference between believing p and not believing p makes a difference to action. Among other things, it means that raising the probability of p does not, on its own, make betting on p more attractive. If you already believe p, getting more evidence for p does not give you more reason to bet on p.[4] Now if we abstract from the agent's situation, it's very hard to see why having a credence cross any threshold (short of 1) should mean that no more evidence matters to p-related decisions. For one thing, it could be that the agent faces a bet on p at astronomical odds, and hence with negative expected utility. But just as importantly, the relation between increased credence and action is continuous. The higher the credence in p, the more bets on p you're prepared to accept. And a marginal increase in the credence of p means a marginal increase in the set of bets on p you're prepared to accept. There aren't any 'jumps' in the function between credence and betting behaviour. The fact that beliefs have a distinctive role in motivating action suggests that the crossing boundary between belief and non-belief should make a significant, or at least a distinctive, difference in actions or dispositions to act. But increasing credence from 0.9785 to 0.9786, or whatever takes you across the threshold, does not make a significant difference.

The Interest-Relative Theory of belief does not face a problem here. Coming to believe that p does have a significant, and distinctive, effect. It means that you're prepared to accept all bets on p that you actually face. The Threshold View implies that there aren't any such nice generalisations about the role of belief in motivating action. That is a serious cost of the view.

3 Thresholds and Closure

The second problem concerns conjunction. It is also set out clearly by Stalnaker.

[4] This is related to the principle that Fantl and McGrath (2009) call 'Safe Reasons', though I'm making a claim here about motivating reasons, not justificatory reasons.

> Reasoning in this way from accepted premises to their deductive consequences (P, also Q, therefore R) does seem perfectly straightforward. Someone may object to one of the premises, or to the validity of the argument, but one could not intelligibly agree that the premises are each acceptable and the argument valid, while objecting to the acceptability of the conclusion. (Stalnaker, 1984, 92)

If categorical belief is having a credence above the threshold, then one can coherently do exactly this. Let x be a number between r and than $r^{1/2}$, such that for an atom of type U has probability x of decaying within a time t, for some t and U. Assume our agent knows this fact, and is faced with two (isolated) atoms of U. Let p be that the first decays within t, and q be that the second decays within t. She should, given her evidence, believe p to degree x, q to degree x, and $p \wedge q$ to degree x^2. If she believed $p \wedge q$ to a degree greater than r, she'd have to either have credences that were not supported by her evidence, or credences that were incoherent. (Or, most likely, both.) So the Threshold View violates the platitude.

This is a well-known argument, so there are many responses to it. The best such response, I think, involves the preface paradox. David Christensen (2005), for example, argues that the preface paradox provides a reason for doubting that beliefs must be closed under entailment, or even must be consistent. Here is his description of the case.

> We are to suppose that an apparently rational person has written a long non-fiction book—say, on history. The body of the book, as is typical, contains a large number of assertions. The author is highly confident in each of these assertions; moreover, she has no hesitation in making them unqualifiedly, and would describe herself (and be described by others) as believing each of the book's many claims. But she knows enough about the difficulties of historical scholarship to realize that it is almost inevitable that at least a few of the claims she makes in the book are mistaken. She modestly acknowledges this in her preface, by saying that she believes the book will be found to contain some errors, and she graciously invites those who discover the errors to set her straight. (Christensen, 2005, 33-4)

Christensen thinks such an author might be rational in every one of her beliefs, even though these are all inconsistent. Although he does not say this, nothing in his discussion suggests that he is using the irrelevance of some of the propositions in the author's defence. So here is an argument that we should abandon closure amongst relevant beliefs.[5]

Christensen's discussion, like many other discussions of the preface paradox, makes frequent use of the fact that examples like these are quite common. We don't have to

[5] As we'll see in the next section, I'm going to agree that a rational agent can have inconsistent beliefs, provided not all of the beliefs in question are relevant to her current interests.

go to fake barn country to find a counterexample to closure. But it seems to me that we need two quite strong idealisations in order to get a real counterexample here.

The first of these is discussed by Ishani Maitra (2010), and is briefly mentioned by Christensen in setting out the problem. We only have a counterexample to closure if the author *believes* every thing she writes in her book. (Indeed, we only have a counterexample if she reasonably believes every one of them. But we'll assume a rational author who only believes what she ought to believe.) This seems unlikely to be true to me. An author of a historical book is like a detective who, when asked to put forward her best guess about what explains the evidence, says "If I had to guess, I'd say . . . " and then launches into spelling out her hypothesis. It seems clear that she need not *believe* the truth of her hypothesis. If she did that, she could not later learn it was true, because you can't learn the truth of something you already believe. And she wouldn't put any effort into investigating alternative suspects. But she can come to learn her hypothesis was true, and it would be rational to investigate other suspects. It seems to me (following here Maitra's discussion) that we should understand scholarly assertions as being governed by the same kind of rules that govern detectives making the kind of speech being contemplated here. And those rules don't require that the speaker believe the things they say without qualification. The picture is that the little prelude the detective explicitly says is implicit in all scholarly work.

There are three objections I know to this picture, none of them particularly conclusive. First, Christensen says that the author doesn't qualify their assertions. But neither does our detective qualify most individual sentences. Second, Christensen says that most people would describe our author as believing her assertions. But it is also natural to describe our detective as believing the things she says in her speech. It's natural to say things like "She thinks it was the butler, with the lead pipe," in reporting her hypothesis. Third, Timothy Williamson (2000) has argued that if speakers don't believe what they say, we won't have an explanation of why Moore's paradoxical sentences, like "The butler did it, but I don't believe the butler did it," are always defective. Whatever the explanation of the paradoxicality of these sentences might be, the alleged requirement that speakers believe what they say can't be it. For our detective cannot properly say "The butler did it, but I don't believe the butler did it" in setting out her hypothesis, even though *believing* the butler did it is not necessary for her to say "The butler did it" in setting out just that hypothesis.[6]

It is plausible that for *some* kinds of books, the author should only say things they believe. This is probably true for travel guides, for example. Interestingly, casual observation suggests that authors of such books are much less likely to write modest prefaces.[7] This makes some sense if those books can only include statements their

[6] I'm assuming here that we should offer a quite different explanation of sentences like *p but I don't believe it* and sentences like *p but I don't know it*. I'll come back to the latter kind of sentence in later chapters. For the reasons given towards the end of Maitra and Weatherson (2010), I don't think that kind of sentence has anything like the significance to the theory of assertion that it is frequently given.

[7] Here's an example from Lonely Planet.

> Things change – prices go up, schedules change, good places go bankrupt. Nothing stays the same. So, if you find things better or worse, recently opened or long-since closed, please tell us and help make the next edition even more accurate and useful.

authors believe, and the authors believe the conjunctions of what they believe.

The second idealisation is stressed by Simon Evnine in his paper "Believing Conjunctions". The following situation does not involve me believing anything inconsistent.

- I believe that what Papi just said, whatever it was, is false.
- Papi just said that the stands at Fenway Park are green.
- I believe that the stands at Fenway Park are green.

If we read the first claim *de dicto*, that I believe that Papi just said something false, then there is no inconsistency. (Unless I also believe that what Papi just said was that the stands in Fenway Park are green.) But if we read it *de re*, that the thing Papi just said is one of the things I believe to be false, then the situation does involve me being inconsistent. The same is true when the author believes that one of the things she says in her book is mistaken. If we understand what she says *de dicto*, there is no contradiction in her beliefs. It has to be understood *de re* before we get a logical problem. And the fact is that most authors do not have *de re* attitudes towards the claims made in their book. Most authors don't even remember everything that's in their books. (I'm not sure I remember how this section started, let alone this book.) Some may argue that authors don't even have the capacity to consider a proposition as long and complicated as the conjunction of all the claims in their book. Christensen considers this objection, but says it isn't a serious problem.

> It is undoubtedly true that ordinary humans cannot entertain book-length conjunctions. But surely, agents who do not share this fairly *superficial* limitation are easily conceived. And it seems just as wrong to say of such agents that they are rationally required to believe in the inerrancy of the books they write. (38: my emphasis)

I'm not sure this is undoubtedly true; it isn't clear that propositions (as opposed to their representations) have lengths. And humans can believe propositions that *can* be represented by sentences as long as books. But even without that point, Christensen is right that there is an idealisation here, since ordinary humans do not know exactly what is in a given book, and hence don't have *de re* attitudes towards the propositions expressed in the book.

I'm actually rather suspicious of the intuition that Christensen is pushing here, that idealising in this way doesn't change intuitions about the case. The preface paradox gets a lot of its (apparent) force from intuitions about what attitude we should have towards real books. Once we make it clear that the real life cases are not relevant to the paradox, I find the intuitions become rather murky. But I won't press this point.

That doesn't quite *say* that everything in the book when it is published, but it strongly implicates that discrepancies between the book and the world as you find it are due to a changing world, not inaccuracies in the book. It certainly isn't the modest preface beloved of fans of the preface paradox.

I believe this kind of comment is standard in a range of Lonely Planet guides, but the particular quote I have here is from Fallon and Yale (2000, 7), a book that I assume is now primarily useful for philosophical uses like this one.

A more important point is that we believers in closure don't think that authors should think their books are inerrant. Rather, we think that authors shouldn't unqualifiedly *believe* the individual statements in their book if they don't believe the conjunction of those statements. Rather, their attitude towards those propositions (or at least some of them) should be that they are probably true.[8] Proponents of the preface paradox know that this is a possible response, and tend to argue that it is impractical. Here is Christensen on this point.

> It is clear that our everyday binary way of talking about beliefs has immense practical advantages over a system which insisted on some more fine-grained reporting of degrees of confidence … At a minimum, talking about people as believing, disbelieving, or withholding belief has at least as much point as do many of the imprecise ways we have of talking about things that can be described more precisely. (96)

Richard Foley makes a similar point.

> There are *deep* reasons for wanting an epistemology of beliefs, reasons that epistemologies of degrees of belief by their very nature cannot possibly accommodate. (Foley, 1993, 170, my emphasis)

It's easy to make too much of this point. It's a lot easier to triage propositions into TRUE, FALSE and NOT SURE and work with those categories than it is to work assign precise numerical probabilities to each proposition. But these are not the only options. Foley's discussion subsequent to the above quote sometimes suggests they are, especially when he contrasts the triage with "indicat[ing] as accurately as I can my degree of confidence in each assertion that I defend." (171) But really it isn't *much* harder to add two more categories, PROBABLY TRUE and PROBABLY FALSE to those three, and work with that five-way division rather than a three-way division. It's not clear that humans as they are actually constructed have a *strong* preference for the three-way over the five-way division, and even if they do, I'm not sure in what sense this is a 'deep' fact about them.[9]

Once we have the five-way division, it is clear what authors should do if they want to respect closure. For any conjunction that they don't believe (i.e. classify as true), they should not believe one of the conjuncts. But of course they can classify every conjunct as probably true, even if they think the conjunction is false, or even certainly false. Still, might it not be considered something of an idealisation to say rational authors must make this five-way distinction amongst propositions they consider? Yes, but it's no more of an idealisation than we need to set up the preface

[8] The position I'm taking here is heavily influenced by the discussion in (Stalnaker, 1984), but I'm not sure whether it's the same in all details. In particular, I don't think it's particularly helpful to describe authors as 'accepting' the propositions they put forward, as Stalnaker does. I'm not sure whether that's because I have a different take on the preface paradox to Stalnaker, or simply that we use 'accept' to pick out slightly different states.

[9] A lot of recent work in experimental philosophy asks subjects to rank propositions on a seven-point scale, from clearly false to clearly true. That's exactly the kind of scale that I think humans easily and frequently construct for themselves. See, among many others, Feltz and Zarpentine (2010).

paradox in the first place. To use the preface paradox to find an example of someone who reasonably violates closure, we need to insist on the following three constraints.

a) They are part of a research community where only asserting propositions you believe is compatible with active scholarship;
b) They know exactly what is in their book, so they are able to believe that one of the propositions in the book is mistaken, where this is understood *de re*; but
c) They are unable to effectively function if they have to effect a five-way, rather than a three-way, division amongst the propositions they consider.

Put more graphically, to motivate the preface paradox we have to think that our inability to have *de re* thoughts about the contents of books is a "superficial constraint", but our preference for working with a three-way rather than a five-way division is a "deep" fact about our cognitive system. Maybe each of these attitudes could be plausible taken on its own (though I'm sceptical of that) but the conjunction seems just absurd.

I'm not entirely sure an agent subject to exactly these constraints is even fully conceivable. (Such an agent is negatively conceivable, in the terminology of Chalmers (2002), but I rather doubt they are positively conceivable.) But even if they are a genuine possibility, why the norms applicable to an agent satisfying that very gerrymandered set of constraints should be considered relevant norms for our state is far from clear. I'd go so far as to say it's clear that the applicability (or otherwise) of a given norm to such an odd agent is no reason whatsoever to say it applies to us. But since the preface paradox only provides a reason for just these kinds of agents to violate closure, we have no reason for ordinary humans to violate closure. So I see no reason here to say that we can have probabilistic coherence without logical coherence, as proponents of the threshold view insist we can have, but which I say we can't have *at least when the propositions involved are salient*. The more pressing question, given the failure of the preface paradox argument, is why I don't endorse a much stronger closure principle, one that drops the restriction to salient propositions. The next section will discuss that point.

4 The Variable Threshold View and Closure

So far I've argued that the Threshold View is committed to denials of multi-premise closure. I've also argued that this is is a very bad feature of the Threshold View, unless it is an unavoidable cost. And I've argued against thinking that the preface paradox provides a reason to resist that argument. But perhaps it is unavoidable. Perhaps, that is, any plausible theory about the relationship between belief and credence will imply that we have to give up on the intuitive multi-premise closure principle. The aim of this section is to show that that's not true. First, I'll argue that we get a reasonably strong closure principle from the Interest Relative Theory of belief. Second, I'll argue that the closure principle we get, although perhaps not as strong as what some theorists want, is as strong a principle as can be justified by arguments for closure.

Here's what we can prove from the Interest Relative Theory of Belief.

Salient Closure: Whenever p and q and their conjunction are all open or salient, and both are believed, and the agent is probabilistically coherent, the agent also believes $p \wedge q$.

Salient Closure is not as strong a principle as some would like, since it is restricted to salient (or open) propositions. But the fact that it is weak isn't a reason to think it is *false*. Indeed, its weakness is a reason to think it might just be true, even if stronger closure principles fail.

The proof of **Salient Closure** is a little complicated, but worth working through. First we'll prove that if the agent believes p, believes q, and p and q are both salient, then the agent prefers believing $p \wedge q$ to not believing it, if $p \wedge q$ is eligible for belief. In what follows $Pr(x|y)$ is the agent's conditional degree of belief in x given y. Since the agent is coherent, we'll assume this is a probability function (hence the name).

1. Since the agent believes that q, they prefer believing that q to not believing that q (by the criteria for belief)
2. So, given p, the agent prefers believing that q to not believing that q. (From 1 and the fact that they believe that p, and that q is salient)
3. So $Pr(q|p) > 1/2$ (from 2)
4. $Pr(q|p) = Pr(p \wedge q|p)$ (by probability calculus)
5. So $Pr(p \wedge q|p) > 1/2$ (from 3, 4)
6. So, if $p \wedge q$ is eligible for belief, then the agent prefers believing that $p \wedge q$ to not believing it, given p (from 5)
7. So, if $p \wedge q$ is eligible for belief, the agent prefers believing that $p \wedge q$ to not believing it (from 6, and the fact that they believe that p, and $p \wedge q$ is salient)

So whenever, p, q and $p \wedge q$ are salient, and the agent believes each conjunct, the agent prefers believing the conjunction $p \wedge q$ to not believing it, if $p \wedge q$ is eligible. Now we have to prove that $p \wedge q$ is eligible for belief, to prove that it is actually believed. That is, we have to prove that (3) follows from (2) and (1), where the initial quantifiers range over actions that are open and salient *tout court*.

(1) $\forall A \forall B \forall r$ ($A \geq_r B \leftrightarrow A \geq_{p \wedge r} B$)
(2) $\forall A \forall B \forall r$ ($A \geq_r B \leftrightarrow A \geq_{q \wedge r} B$)
(3) $\forall A \forall B \forall r$ ($A \geq_r B \leftrightarrow A \geq_{p \wedge q \wedge r} B$)

Assume that (3) isn't true. That is, there are A, B and s such that $\neg(A \geq_s B \leftrightarrow A \geq_{p \wedge q \wedge s} B)$. By hypothesis s is active, and consistent with $p \wedge q$. So it is the conjunction of relevant, salient propositions. Since q is salient, this means $q \wedge s$ is also active. Since s is consistent with $p \wedge q$, it follows that $q \wedge s$ is consistent with p. So $q \wedge s$ is a possible substitution instance for r in (1). Since (1) is true, it follows that $A \geq_{q \wedge s} B \leftrightarrow A \geq_{p \wedge q \wedge s} B$. By similar reasoning, it follows that s is a permissible substitution instance in (2), giving us $A \geq_s B \leftrightarrow A \geq_{q \wedge s} B$. Putting the last two biconditionals together we get $A \geq_s B \leftrightarrow A \geq_{p \wedge q \wedge s} B$, contradicting our hypothesis that there is a

counterexample to (3). So whenever (1) and (2) are true, (3) is true as well, assuming p, q and $p \wedge q$ are all salient.

So the Interest Relative Theory of belief does give us a closure principle, namely **Salient Closure**. But maybe we should not be happy with that. Some theorists will surely object that this isn't enough of a closure principle, since closure should not be restricted in this way. Here I want to defend the plausibility of the qualification. Let's start with what I take to be the most important argument for closure, the passage from Stalnaker's *Inquiry* that I quoted above.

> Reasoning in this way from accepted premises to their deductive consequences (P, also Q, therefore R) does seem perfectly straightforward. Someone may object to one of the premises, or to the validity of the argument, but one could not intelligibly agree that the premises are each acceptable and the argument valid, while objecting to the acceptability of the conclusion. (Stalnaker, 1984, 92)

Stalnaker's wording here is typically careful. The relevant question isn't whether we can accept p, accept q, accept p and q entail r, and reject r. As Christensen (2005, Ch. 4) notes, this is impossible even on the threshold view, as long as the threshold is above 2/3. The real question is whether we can accept p, accept q, accept p and q entail r, and *fail* to accept r. And this is always a live possibility on any threshold view, though it seems absurd at first that this could be coherent.

But it's important to note how *active* the verbs in Stalnaker's description are. When faced with a valid argument we have to *object* to one of the premises, or the validity of the argument. What we can't do is *agree* to the premises and the validity of the argument, while *objecting* to the conclusion. I agree. If we are really *agreeing* to some propositions, and *objecting* to others, then all those propositions are salient. And in that case closure, deductive coherence, is mandatory. This doesn't tell us what we have to do if we haven't previously made the propositions salient in the first place.

The position I endorse here is very similar in its conclusions to that endorsed by Gilbert Harman in *Change in View*. There Harman endorses the following principle. (At least he endorses it as true – he doesn't seem to think it is particularly explanatory because it is a special case of a more general interesting principle.)

> **Recognized Logical Implication Principle** One has reason to believe P if one *recognizes* that P is logically implied by one's view. (Harman, 1986, 17)

This seems right to me, both what it says and its implicature that the reason in question is not a conclusive reason. My main objection to those who use the preface paradox to argue against closure is that they give us a mistaken picture of what we have *to do* epistemically. When I have inconsistent beliefs, or I don't believe some consequence of my beliefs, that is something I have a reason to deal with at some stage, something I have to do. When we say that we have things to do, we don't mean that we have to do them *right now*, or instead of everything else. My current list of things to do includes cleaning my bathroom, yet here I am writing this paper, and (given the relevant deadlines) rightly so. We can have the job of cleaning up our

epistemic house as something to do while recognising that we can quite rightly do other things first. But it's a serious mistake to infer from the permissibility of doing other things that cleaning up our epistemic house (or our bathroom) isn't something to be done. The bathroom won't clean itself after all, and eventually this becomes a problem.

There is a possible complication when it comes to tasks that are very low priority. I used to own a house with a typically cluttered attic. I used to think that the attic could well be cleaned, or at least it could be cleaner, but there are no imaginable circumstances under which something else wouldn't be higher priority. Given that, should I really have *clean the attic* on the list of things to be done? Similarly, there might be implications I haven't followed through that it couldn't possibly be worth my time to sort out. Are they things to be done? I think it's worthwhile recording them as such, because otherwise we might miss opportunities to deal with them in the process of doing something else. I don't need to put off anything else in order to clean the attic, but if I'm up there for independent reasons I should bring down some of the garbage. Similarly, I don't need to follow through implications mostly irrelevant to my interests, but if those propositions come up for independent reasons, I should deal with the fact that some things I believe imply something I don't believe. Having it be the case that all implications from things we believe to things we don't believe constitute jobs to do (possibly in the loose sense that cleaning my attic is something to do) has the right implications for what epistemic duties we do and don't have.[10]

While waxing metaphorical, it seems time to pull out a rather helpful metaphor that Gilbert Ryle develops in *The Concept of Mind* at a point where he's covering what we'd now call the inference/implication distinction. (This is a large theme of chapter 9, see particularly pages 292-309.) Ryle's point in these passages, as it frequently is throughout the book, is to stress that minds are fundamentally active, and the activity of a mind cannot be easily recovered from its end state. Although Ryle doesn't use this language, his point is that we shouldn't confuse the difficult activity of drawing inferences with the smoothness and precision of a logical implication. The language Ryle does use is more picturesque. He compares the easy work a farmer does when sauntering down a path from the hard work he did when building the path. A good argument, in philosophy or mathematics or elsewhere, is like a well made path that permits sauntering from the start to finish without undue strain. But from that it doesn't follow that the task of coming up with that argument, of building that path in Ryle's metaphor, was easy work. The easiest paths to walk are often the hardest to build. Path-building, smoothing out our beliefs so they are consistent and closed under implication, is hard work, even when the finished results look clean and straightforward. Its work that we shouldn't do unless we need to. But making sure our beliefs are closed under entailment even with respect to irrelevant propositions is suspiciously like the activity of buildings paths between points without first checking you need to walk between them.

[10]This example feels a little dated, since I wrote it when living in a spacious house up in Ithaca, not an apartment in Manhattan. But the thing about apartments is that you can't leave space in them poorly used for too long and not have it become a pressing job soon enough. No matter; it turns out that when you sell a house you don't sell the right to use it as a metaphor.

For a less metaphorical reason for doubting the wisdom of this unchecked commitment to closure, we might notice the difficulties theorists tend to get into all sorts of difficulties. Consider, for example, the view put forward by Mark Kaplan in *Decision Theory as Philosophy*. Here is his definition of belief.

> You count as believing P just if, were your sole aim to assert the truth (as it pertains to P), and you only options were to assert that P, assert that ¬P or make neither assertion, you would prefer to assert that P. (109)

Kaplan notes that conditional definitions like this are prone to Shope's conditional fallacy. If my sole aim were to assert the truth, I might have different beliefs to what I now have. He addresses one version of this objection (namely that it appears to imply that everyone believes their sole desire is to assert the truth) but as we'll see presently he can't avoid all versions of it.

These arguments are making me thirsty. I'd like a beer. Or at least I think I would. But wait! On Kaplan's theory I can't think that I'd like a beer, for if my sole aim were to assert the truth as it pertains to my beer-desires, I wouldn't have beer desires. And then I'd prefer to assert that I wouldn't like a beer, I'd merely like to assert the truth as it pertains to my beer desires.

Even bracketing this concern, Kaplan ends up being committed to the view that I can (coherently!) believe that p even while regarding p as highly improbable. This looks like a refutation of the view to me, but Kaplan accepts it with some equanimity. He has two primary reasons for saying we should live with this. First, he says that it only looks like an absurd consequence if we are committed to the Threshold View. To this all I can say is that *I* don't believe the Threshold View, but it still seems absurd to me. Second, he says that any view is going to have to be revisionary to some extent, because our ordinary concept of belief is not "coherent" (142). His view is that, "Our ordinary notion of belief both construes belief as a state of confidence short of certainty and takes consistency of belief to be something that is at least possible and, perhaps, even desirable" and this is impossible. I think the view here interprets belief as a state less than confidence and allows for as much consistency as the folk view does (i.e. consistency amongst salient propositions), so this defence is unsuccessful as well.

5 Psychological Evidence that Stakes Matter to Belief

3

How Interest-Relativity Affects Epistemology

4

Against the Threshold View of Belief

5

Fantl and McGrath

1 Fantl and McGrath on Interest-Relativity

2 Fantl and McGrath's Argument that Not Only Belief is Interest-Relative

Fantl and McGrath's primary complaint against the Interest Relative Theory of Belief is that it is not strong enough to entail principles such as (JJ).

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

It's true that the Interest Relative Theory cannot be used to derive (JJ), at least on its intended reading. But that's because on the intended reading, it is false, and the Interest Relative Theory is true. So the fact that (JJ) can't be derived is a feature, not a bug.

The problem arises because of cases like the case we discussed at the end of chapter 1, namely cases where the agent has irrational beliefs elsewhere in her web of belief. Let's lay out the case carefully, because it will have a large role in the rest of this chapter. Here's what we're going to stipulate about S.

- She knows that p and q are independent, so her credence in any conjunction where one conjunct is a member of $\{p, \neg p\}$ and the other is a member of $\{q, \neg q\}$ will be the product of her credences in the conjuncts.
- Her credence in p is 0.99, just as the evidence supports.
- Her credence in q is also 0.99. This is unfortunate, since the rational credence in q given her evidence is 0.01.
- She has a choice between taking and declining a bet with the following payoff structure.[1]
  - If $p \wedge q$, she wins \$100.
  - If $p \wedge \neg q$, she wins \$1.

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

  - If $\neg p$, she loses \$1000.

- The marginal utility of money is close enough to constant that expected dollar returns correlate more or less precisely with expected utility returns.

As can be easily computed, the expected utility of taking the bet given her credences is positive, it is just over \$89. Our agent *S* takes the bet. She doesn't compute the expected utility, but she is sensitive to it.[2] That is, had the expected utility given her credences been close to 0, she would have not acted until she made a computation. But from her perspective this looks like basically a free \$100, so she takes it. Happily, this all turns out well enough, since p is true. But it was a dumb thing to do. The expected utility of taking the bet given her evidence is negative, it is a little under -\$8. So she isn't warranted, given her evidence, in taking the bet.

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

1. p is not justified enough to warrant her in taking the bet.
2. She believes p.[3]
3. This belief is rational.

The argument for 1 is straightforward. She isn't warranted in taking the bet, so p isn't sufficiently warranted to justify it. This is despite the fact that p is obviously relevant. Indeed, given p, taking the bet strictly dominates declining it. But still, p doesn't warrant taking this bet, because nothing warrants taking a bet with negative expected utility. Had the rational credence in p been higher, then the bet would have been reasonable. Had the reasonable credence in p been, say, 0.9999, then she would have been reasonable in taking the bet, and using p as a reason to do so. So there's a good sense in which p simply isn't warranted enough to justify taking the bet.[4]

The argument for 2 is that she has a very high credence in p, this credence is grounded in the evidence in the right way, and it leads her to act as if p is true, e.g. by taking the bet. It's true that her credence in p is not 1, and if you think credence 1 is needed for belief, then you won't like this example. But if you think that, you won't think there's much connection between (JJ) and pragmatic conditions in epistemology either. So that's hardly a position a defender of Fantl and McGrath's position can hold.[5]

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

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

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

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

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

I don't think the argument here strictly needs the assumption I'm about to make, but I think it's helpful to see one very clear way to support the argument of the last paragraph. The working assumption of this book is that talking about beliefs and talking about credences are simply two ways of modelling the very same things, namely minds.[6] If the agent both has a credence 0.99 in p, and believes that p, these are not two different states. Rather, there is one state of the agent, and two different ways of modelling it. So it is implausible, if not incoherent, to apply different valuations to the state depending on which modelling tools we choose to use. That is, it's implausible to say that while we're modelling the agent with credences, the state is rational, but when we change tools, and start using beliefs, the state is irrational. Given this outlook on beliefs and credences, premise 3 seems to follow immediately from the setup of the example.

3 Objections and Replies

So that's the argument that (JJ) is false. And if it's false, the fact that the Interest Relative Theory doesn't entail it is a feature, not a bug. But there are a number of possible objections to that position. I'll spend the rest of this chapter going over them.[7]

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

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

C. So, S does not justifiably believe that p

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

It might be controversial to use this very case to reject premise 1. But the falsity of premise 1 should be clear on independent grounds. What p entails is that S will be best off by taking the bet. But there are lots of things that will make me better off that I shouldn't do. Imagine I'm standing by a roulette wheel, and the thing that

[6] Recall the discusison in section 2.1 of just this point.

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

will make me best off is betting heavily on the number than will actually come up. It doesn't follow that I should do that. Indeed, I should not do it. I shouldn't place any bets at all, since all the bets have a highly negative expected return.

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

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

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

It may seem better to use p combined with the fact that conditional on p, taking the bet has the highest *expected* utility. But actually that's still not much of a reason to take the bet. Think again about cases, completely normal cases, where the action with the best outcome is an *ex ante* unjustifiable risk. Call that action ϕ, and let $B\phi$ be the proposition that ϕ has the best outcome. Then $B\phi$ is true, and conditional on $B\phi$, ϕ has an excellent expected return. But doing ϕ is still running a dumb risk. Since these kinds of cases are normal, it seems it will very often be the case that this form of inference leads from truth to falsity. So it's not a reliable inductive inference.

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

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

Reply: It's possible at this point that we get to a dialectical impasse. I think this inference is non-deductive, because I think the example we're discussing here is one

where the premises are true and the conclusion false. Presumably someone who doesn't like the example will think that it is a good deductive inference.

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

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

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

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

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

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

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

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

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

Objection: Since the ideal agent in S's position would not believe p, it follows that there is no *propositional* justification for p. Moreover, doxastic justification requires propositional justification[8] So S is not doxastically justified in believing p. That is, she isn't justified in believing p.

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

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

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

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

Objection: If S is justified in believing p, then S can use p as a premise in practical reasoning. If S can use p as a premise in practical reasoning, and p is true, and her belief in p is not Gettiered, then she knows p. By hypothesis, her belief is true, and her belief is not Gettiered. So she should know p. But in the previous section it was argued that she doesn't know p. So by several steps of modus tollens, she isn't justified in believing p.[9]

Reply: Like the previous objection, this one turns on an equivocation, this time over the neologism 'Gettiered'. Some epistemologists use this to simply mean that a belief is justified and true without constituting knowledge. By that standard, the third sentence is false. Or, at least, we haven't been given any reason to think that it is true. Given everything else that's said, the third sentence is a raw assertion that S knows that p, and I don't think we should accept that.

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

But it's worth pausing for a bit to reflect on why we should say that S does not know that p. My main reason for saying this comes from thinking about how we lay out examples in decision theory. Compare the following two examples.

Case 1 S has a counter and two options about what to do with it. If she places it on the table, she will get \$100,000. If she places it in her pocket, she will get nothing. There are no other salient choices, and there are no other effects associated with her action, and she prefers more money to less. What should she do?

Case 2 S has a counter and two options about what to do with it. If she places it on the table, she is betting \$100,000 that the ball in the roulette wheel will land on a red number. There is a $^{18}/_{37}$ probability that this will happen, and if she bets, she'll win \$100,000 if the ball lands on a red, and lose \$100,000 if it does not.

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

> If she puts the counter in her pocket, she'll get nothing. There are no other salient choices, and there are no other effects associated with her action, and she prefers more money to less, and the marginal utility of money is constant for her over the financial range being discussed. What should she do?

If we were in a standard decision theory class, the answers would be easy. She should put the counter on the table in the first case, and in her pocket in the second.

But there's nothing in the description of the two cases that rules out their describing the very same facts. If in Case 2 the ball does actually land on a red, then everything that's true in Case 1 will be true in Case 2. And in Case 1, she should put the counter on the table. Presumably it would be a mistake to infer by modus tollens that the ball will not land on a red in Case 2, although a literal reading of the last few paragraphs would suggest that would be a good inference to draw. What is going on?

I think the obvious thing to say is that when we set up cases like Case 1 and Case 2, we don't just stipulate what's true in the story. We also stipulate that the agent stands in some epistemic relationship (broadly construed) to the facts as laid out in the story. So even if the ball in the roulette wheel in Case 2 lands on a red number, that doesn't turn the case into Case 1. That's because for us to be in Case 1, the ball would not only have to land on a red, but *S* would have to stand in the appropriate epistemic relation to that landing.

The big question then is what this salient epistemic relation might happen to be. I think it is knowledge. I don't have anything like a *conclusive* argument for this. But I think there are a couple of reasons that push us that way. If we go with a relation weaker than knowledge, such as justified belief, then we won't be able to explain the fact that agents are meant to be able to rely on the structure of the problem, even in surprising circumstances. Agents are meant to have an indefeasible belief in the structure of the problem as stated, which is hard to do make consistent with the rationality of the agents unless you suppose that they know the problem has that structure. If you go with relation much stronger than knowledge, such as certainty, you'll be left with the problem that decision theory is not a particularly useful tool for modelling real-world interactions, since it is hard to ever be certain of anything. More positively, it's very natural when describing a decision theoretic problem to say that the agent knows various things that are in the setting out of the problem.

Now these problems don't only show up in decision theory textbooks. They also show up in, for example, epistemology papers. The problem with which we opened this section is just such a problem. Now let's say that the standard for including something in the description of the problem is knowledge. Then we could have simplified the problem a lot, as follows.

- If *S* takes the bet, she wins \$100 if q and \$1 if $\neg q$, and will not lose any money.
- If *S* declines the bet, she gets nothing.

And given that description of the bet, *S* should obviously take it, whatever her views on q. Since *S* shouldn't take the bet, that means the bet has been misdescribed. And that means *S* does not know that p.

6

Hawthorne and Stanley

7

Objections to Interest Relativity

1 Temporal Embeddings

Michael Blome-Tillmann has argued that his version of Lewisian contextualism is preferable to Interest Relative Invariantism (Blome-Tillmann, 2009a)[1] We'll discuss one argument he gives, an argument from temporal embeddings, in this section, and a distinct argument involving conjunctions in the next section.

Blome-Tillmann's argument usees a variant of the well-known bank cases.[2] 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 (7).

(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 Interest Relative Invariantism. As Blome-Tillmann notes, it isn't really a situation where Hannah's stakes change. She was never actually in a high stakes situation. At most her perception of her stakes change; she thought she was in a high-stakes situation, then realised that she wasn't. Blome-Tillmann argues that even this change in perceived stakes can be enough to make (1) true if Interest Relative Invariantism 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 Interest Relative Invariantism.

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 Interest Relative Invariantism. Any theory should say that if a person is worrying about $\neg O$ possibilities, and planning around them, they don't know O. If Hannah is rational, that will describe her on Thursday, but not on Friday.

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

[2] See Stanley (2005a) for the original versions that Blome-Tillman is building on.

So (7) is true not because Hannah's practical situation changes between Thursday and Friday, but because her psychological state changes.

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

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

But it's worth remembering that if Hannah does irrationally ignore $\neg O$ possibilities, she is being irrational with respect to O. And it's very plausible that this irrationality defeats knowledge. That is, you can't be irrational with respect to a proposition and know it. Irrationality excludes knowledge. That may look a little like a platitude, so it's worth spending a little time on how it can lead to some quirky results for reasons independent of Interest Relative Invariantism.[3]

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

I think Bobby's case is just like Hannah's, at least under the assumption that Hannah simply ignores the significance of O to her practical deliberation. In both cases, defective mental states elsewhere in their cognitive architecture defeat knowledge claims. And in that kind of case, we should expect sentences like (1) to be true, even if they appear counterintuitive before we've worked through the details. The crucial point is that once we work through the details, we see that somewhat distant changes in the rest of the cognitive system changes what the agent knows. So, a little counterintuitively, it can be the case that an agent knows something after a distant change in the system, but not before. That's all that happens in Hannah's case.

[3]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.

2 Conjunctions

Blome-Tillman also argues that Interest Relative Invariantism is committed to certain kinds of counterintuitive conjunctions. The form of the conjunctions are that two people differ in what they know, although the two are alike in some respects relevant to knowledge. The short version of what I'm going to argue is that these conjunctions are always acceptable unless we have good reason to believe the two are alike in *every* respect relevant to knowledge. But before we get to the examples, I want to note some points about 'enough'. Let's start with an example unconnected to knowledge. 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 an argument Blome-Tillmann (2009a, ??) makes concerning knowledge and the idea of having enough evidence to know. Here is how he puts the argument. (I've changed the numbering and some terminology for consistency with this chapter.)

> 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 Interest Relative Invariantism is the view that whether one knows p depends on one's practical situation, Interest Relative Invariantism 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 Interest Relative Invariantism 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, Interest Relative Invariantism 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.' (Blome-Tillmann, 2009a, ??)

The 'reformulation' is obviously bad as a reformulation, since having enough evidence to know p isn't the same as being in a position to know it. To see this, note that having enough money to buy the computer is not the same as being in a position to buy it. George has enough money to buy the computer, but he isn't in a position to buy it.

But set that aside. If the original argument was bad, perhaps this isn't simply a reformulation of a bad argument, but a different and better argument against Interest Relative Invariantism. The argument would be that Interest Relative Invariantism 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.

This objection, however, fails on the simple ground that (6) is true. Or, at least, there is no reason to believe that (6) is false. The argument Blome-Tillman makes seems simply to be an appeal to the unintuitiveness of (6). But that appeal loses its force if we have good theoretical reasons to think that sentences like (6) can be true. And indeed we have such good theoretical reasons, reasons that are independent of Interest Relative Invariantism.

Any epistemological theory that denies that what one is in a position to know supervenes on one's evidence will allow that sentences like (6) can be true.[4] 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 (6) can be true. Now I think any particular claim about the existence of a defeater that

[4]The supervenience claim here might seem insanely strong if we take it to be a *modal* supervenience claim. But since John and Paul are in the same world, presumably that's more than Blome-Tillman is really committed to. He's only committed to the claim that any two agents *in the same world* that have the same evidence have the same knowledge. So we can't argue against the presupposed supervenience claim by considering cases where one person knows that p while a modal counterpart of theirs has the same evidence although p is false.

doesn't supervene on evidence will be controversial. But there are so many different kinds of candidates that it should be obvious that there are some such candidates. Here are three possibilities; I hope that each reader finds at least one persuasive.

> *Logic and the Oracle*
> Graham, Crispin and Ringo have an audience with the Delphic Oracle, and they are told $\neg p \vee q$ and $\neg\neg p$. Graham is a relevant logician, so if he inferred $p \wedge q$ from these pronouncements, his belief in the invalidity of disjunctive syllogism would be a doxastic defeater, and the inference would not constitute knowledge. Crispin is an intuitionist logician, so if he inferred $p \wedge q$ from these pronouncements his belief in the invalidity of double negation elimination would be a doxastic defeater, and the inference would not constitute knowledge. Ringo has no deep views on the nature of logic. Moreover, in the world of the story classical logic is correct. So if Ringo were to infer $p \wedge q$ from these pronouncements, his belief would constitute knowledge. Now Graham's and Crispin's false beliefs about entailment are not $p \wedge q$-relevant evidence, so all three of them have the same $p \wedge q$-relevant evidence, but only Ringo is in a position to know $p \wedge q$.
>
> *Missing iPhone* (after Harman (1973)).
> Lou and Andy both get evidence E, which is strong inductive evidence for p. If Lou were to infer p from E, his belief would constitute knowledge. Andy has just missed a phone call from a trusted friend. The friend left a voicemail saying $\neg p$, but Andy hasn't heard this yet. If Andy were to infer p from E, his friend's phone call and voicemail would constitute defeaters, so he wouldn't know p. But phone calls and voicemails you haven't got aren't evidence you have. So Lou and Andy have the same p-relevant evidence, but only Lou is in a position to know p.
>
> *Fake Barns and Motorcycles* (after Gendler and Hawthorne (2005))
> Bob and Levon are travelling through Fake Barn Country. Bob is on a motorcycle, Levon is on foot. They are in an area where the barns are, surprisingly, real for a little ways around. On his motorcycle, Bob will soon come to fake barns, but Levon won't hit any fakes for a long time. They are both looking at the same real barn. If Bob inferred it was a real barn, not a fake, the fakes he is speeding towards would be defeaters. But Levon couldn't walk that far, so those barns don't defeat him. So Bob and Levon have the same evidence, but only Levon is in a position to know that the barn is real.

I'm actually not sure what plausible theory would imply that what different agents are in a position to know depends on *nothing* except for what evidence they have. The only theory that I can imagine with that consequence is the conjunction of evidentialism about justification and a justified true belief theory of knowledge. So

really there's no reason to think that implying sentences like (6) is a mark against a theory.

It's been suggested to me[5] that there are other more problematic conjunctions in the neighbourhood. For instance, we might worry that Interest Relative Invariantism implies that (7) is true.

(7) John and Paul are alike in every respect relevant to knowledge of p, but John is in a position to know p, and Paul isn't.

That would be problematic, but there's no reason to think Interest Relative Invariantism implies it. Indeed, Interest Relative Invariantism entails that the first conjunct is false, since John and Paul are unlike in one respect that Interest Relative Invariantism loudly insists is relevant. Perhaps we can do better with (8).

(8) John and Paul are alike in every respect relevant to knowledge of p except their practical interests, but John is in a position to know p, and Paul isn't.

That is something Interest Relative Invariantism implies, but it seems more than a little question-begging to use its alleged counterintuitiveness against Interest Relative Invariantism. After all, it's simply a statement of Interest Relative Invariantism itself. If someone had alleged that Interest Relative Invariantism should be accepted because it was so intuitive, I guess noting how odd (8) looks would be a response to them. But that's not the way Interest Relative Invariantism has been defended here.[6] I've defended it by noting what a good job it does of handling difficult puzzles to do with the role of credences in philosophy of mind and epistemology. If the outcome is a little counterintuitive, that's not too surprising. It's par for the course when solving difficult puzzles.

3 Knowledge by Indifference and by Wealth

Gillian Russell and John Doris (2009) argue that Jason Stanley's account of knowledge leads to some implausible attributions of knowledge. Insofar as my theory agrees with Stanley's about the kinds of cases they are worried about, their objections are also objections to my theory. I'm going to argue that Russell and Doris's objections turn on principles that are *prima facie* rather plausible, but which ultimately we can reject for independent reasons.[7]

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

[5] By both Jeremy Fantl and Juan Comesaña.

[6] And for what it's worth, I don't think it's how it is defended by others in the literature either.

[7] 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.

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, ??). 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, ??) 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, ??)

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. I've been proposing a model where practical considerations enter debates about knowledge through two main channels: through the definition of belief, and through distinctive kinds of defeaters. It seems the second channel is particularly relevant here. And we have, somewhat surprisingly, independent evidence to think that indifference and wealth do matter to defeaters.

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

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

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

her wealth she knows things she wouldn't otherwise know, even before her money goes to work. Again, her money doesn't generate knowledge - - the original reliable report is the knowledge generator - but her wealth means that a would-be defeater doesn't gain traction.

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

4 Stakes and Odds

In a so far unpublished note, Mark Schroeder (2008) has argued that interest-relative invariantists have erred by stressing variation in stakes as being relevant to knowledge. He argues, using examples of forced choice, that what is really relevant is the odds at which the agent has to make bets. Of course due to the declining marginal utility of material goods, high stakes bets will often be long odds bets. So there's a correlation between stakes and odds. But when the correlation comes apart, Schroeder argues convincingly that it's the odds and not the stakes that are relevant to knowledge.

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

Although I think the variable threshold view gets those cases right, I don't take those examples to be a crucial part of the argument for the view. The core argument is that the view provides a better answer to Sturgeon's challenge of how we should integrate credences and beliefs into a single model. If it turned out that the facts about the examples were less clear than we thought, that wouldn't *undermine* the argument for the variable threshold view, since those facts weren't part of the original argument. But if it turned out that the facts about those examples were quite different to what the variable threshold view 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 (2010). 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. 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, ??)

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 the variable threshold view predicts. The change isn't trivial, even if due to the smallish sample size it isn't statistically significant in this sample. But should the variable threshold view have predicted a larger change? To figure this out, we need to ask three questions.

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

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

Feltz and Zarpentine say that the costs in"High Stakes Bridge [are] very costly—certain death—whereas the costs in Low Stakes Bridge are likely some minor injuries and embarrassment." (Feltz and Zarpentine, 2010, ??) 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.[8] 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.)

[8] 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.

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.

But the big issue is the third question. John has a lot of information that the bridge will support his truck. If I've tested something for sturdiness two or three times, and it has worked, I won't even think about testing it again. Consider what evidence you need before you'll happily stand on a particular chair to reach something in the kitchen, or put a heavy television on a stand. Supporting a weight is the kind of thing that either fails the first time, or works fairly reliably. Obviously there could be some strain-induced effects that cause a subsequent failure[9], but John really has a lot of evidence that the bridge will support him.

Given those three answers, it seems to me that it is a reasonable bet to cross the bridge. At the very least, it's no more of an unreasonable bet than the bet I make every day crossing a busy highway by foot. So I'm not surprised that 64% of the subjects agreed that John knew the bridge would hold him. At the very least, that result is perfectly consistent with the variable threshold view, 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 the variable threshold view if the subjects are reliable. I can think of two reasons why they might not be. First, subjects tend to massively discount the costs and likelihoods of traffic related injuries. In most of the country, the risk of death or serious injury through motor vehicle accident is much higher than the risk of death or serious injury through some kind of crime or other attack, yet most people do much less to prevent vehicles harming them than they do to prevent criminals or other attackers harming them.[10] Second, only 73% of this subjects in *this very experiment* said that John knows the bridge will support him in **Low Stakes Bridge**. This is just absurd. Unless the subjects endorse an implausible kind of scepticism, something has gone wrong with the experimental design. Given the fact that the experiment points broadly in the direction of the theory I favour, and that with some plausible assumptions it is perfectly consistent with that theory, and the unreliability of the subjects, I don't think this kind of experimental work threatens the variable threshold view.

5 Neta

6 DeRose

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

[10] 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

Contextualist and Relativist Alternatives

1 Contextualism and Questions

2 Stanley and Blome-Tillman on Modal Embeddings

Jason Stanley had argued that the fact that Interest Relative Invariantism (hereafter, IRI) has counterintuitive consequences when it comes to knowledge ascriptions in modal contexts shouldn't count too heavily against IRI, because contextualist approaches are similarly counterintuitive. In particular, he argues that the theory that 'knows' is a contextually sensitive quantifier, plus the account of quantifier-domain restriction that he developed with Zoltán Gendler Szabó (Stanley and Szabó, 2000), has false implications when it is applied to knowledge ascriptions in counterfactuals. Michael Blome-Tillmann (2009a) disagrees, but I don't think he provides very good reasons for disagreeing. In fact, Stanley's argument seems to be a very good argument against Lewisian contextualism about 'knows'.[1] Let's start by reviewing how we got to this point.

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

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

We naturally interpret (1) the first way, and (2) the second way.

(1) If I had won the last Presidential election, everyone who voted for me would regret it by now.
(2) If Hilary Clinton had been the Democratic nominee in the last Presidential election, everyone who voted for Barack Obama would have voted for her.

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

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

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

Surprisingly, it's hard to get the second of these readings. Or, at least, it is hard to *avoid* the availability of the first reading. Typically, if we restrict our attention to the Cs, then when we embed the quantifier in the consequent of a counterfactual, the restriction is to the things that would be C, not to the actual Cs.[2]

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

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

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

(3) If there were no philosophers, then the philosophers doing research in the field of applied ethics would be missed most painfully by the public.
(4) If there were no beer, everybody drinking beer on a regular basis would be much healthier.
(5) If I suddenly were the only person alive, I would miss the Frege scholars most.

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

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

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

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

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

- I'm engaged in sceptical doubts;
- there is beer in the fridge
- I've forgotten what's in the fridge; and
- I've got normal vision, so if I check the fridge I'll see what's in it.

In that context it seems (6) is false, since it would only be true if Cartesian doubts weren't salient.

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

But the only way to get that to come out false, and false for the right reasons, is to fix on which worlds we're actually ignoring (i.e., include in the quantifier domain worlds where I'm the victim of an evil demon), but look at worlds that would be ruled out with the counterfactually available evidence. We don't want the sentence to be false because I've actually forgotten what's in the fridge. And we don't want it to be true because I would be ignoring Cartesian possibilities. In the terminology above, we would need *If p were true, all Fs would be Gs* to mean *If p were true, then every actual C that were F would also be G*. We haven't got any reason yet to think that's even a *possible* disambiguation of (6).

But let's make things easy for the contextualist and assume that it is. Stanley's point is that the contextualist needs even more than this. They need it to be by far the *preferred* disambiguation, since in the context I describe the natural reading of (6) (given sceptical intuitions) is that it is false because my looking in the fridge wouldn't rule out Cartesian doubts. And they need it to be the preferred reading even though there are alternative readings that are (a) easier to describe, (b) of a kind more commonly found, and (c) true. Every principle of contextual disambiguation we have pushes us away from thinking this is the preferred disambiguation. This is the deeper challenge Stanley raises for contextualists, and it hasn't yet been solved.

3 Motivations for Relativism

Bibliography

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

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

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

Chalmers, David. 2002. "Does Conceivability Entail Possibility?" In Tamar Szabó Gendler and John Hawthorne (eds.), *Conceivability and Possibility*, 145–200. Oxford: Oxford University Press.

Christensen, David. 2005. *Putting Logic in Its Place*. Oxford: Oxford University Press.

Coates, John. 1996. *The Claims of Common Sense*. Cambridge: Cambridge University Press.

Cohen, L. Jonathan. 1989. "Belief and Acceptance." *Mind* 98:367–389, doi:10.1093/mind/XCVIII.391.367.

Cohen, Stewart. 1999. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives* 13:57–89, doi:10.1111/0029-4624.33.s13.3.

DeRose, Keith. 2009. *The Case for Contextualism: Knowledge, Skepticism and Context*. Oxford: Oxford.

Evnine, Simon. 1999. "Believing Conjunctions." *Synthese* 118:201–227, doi:10.1023/A:1005114419965.

Fallon, Steve and Yale, Pat. 2000. *London*. Melbourne: Lonely Planet Publications, second edition.

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

Fara, Delia Graff. 2000. "Shifting Sands: An Interest-Relative Theory of Vagueness." *Philosophical Topics* 28:45–81. This paper was published under the name 'Delia Graff'.

Feltz, Adam and Zarpentine, Chris. 2010. "Do You Know More When It Matters Less?" *Philosophical Psychology* Retrieved from http://faculty.schreiner.edu/adfeltz/Papers/Know%20more.pdf.

Foley, Richard. 1993. *Working Without a Net*. Oxford: Oxford University Press.

Gendler, Tamar Szabó and Hawthorne, John. 2005. "The Real Guide to Fake Barns: A Catalogue of Gifts for Your Epistemic Enemies." *Philosophical Studies* 124:331–352, doi:10.1007/s11098-005-7779-8.

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

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

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

—. 1986. *Change in View*. Cambridge, MA: Bradford.

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.

Hunter, Daniel. 1996. "On the Relation Between Categorical and Probabilistic Belief." *Noûs* 30:75–98, doi:10.2307/2216304.

Kaplan, Mark. 1996. *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.

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

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

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.

Lewis, David. 1993. "Many, But Almost One." In Keith Campbell, John Bacon, and Lloyd Reinhardt (eds.), *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*, 23–38. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511625343.010. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 164-182.

—. 1994. "Reduction of Mind." In Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, 412–431. Oxford: Blackwell, doi:10.1017/CBO9780511625343.019. Reprinted in *Papers in Metaphysics and Epistemology*, pp. 291-324.

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

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

Machina, Kenton. 1976. "Truth, Belief and Vagueness." *Journal of Philosophical Logic* 5:47–78, doi:10.1007/BF00263657.

Maitra, Ishani. 2010. "Assertion, Norms and Games."

Maitra, Ishani and Weatherson, Brian. 2010. "Assertion, Knowledge and Action." *Philosophical Studies* , doi:10.1007/s11098-010-9542-z.

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.

Portmore, Douglas. 2009. "Consequentializing." *Philosophy Compass* 4:329–347, doi:10.1111/j.1747-9991.2009.00198.x.

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

Ryle, Gilbert. 1949. *The Concept of Mind*. New York: Barnes and Noble.

Schroeder, Mark. 2008. "Stakes, Withholding and the Subject-Sensitivity of 'Knows'." Retrieved from `http://www-rcf.usc.edu/~maschroe/research/Schroeder_Stakes_and_Withholding.pdf`.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

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

—. 2005b. *Knowledge and Practical Interests*. Oxford: Oxford University Press.

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

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

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

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

Weatherson, Brian. 2002. "Keynes, Uncertainty and Interest Rates." *Cambridge Journal of Economics* 26:47–62, doi:10.1093/cje/26.1.47.

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

—. 2005b. "True, Truer, Truest." *Philosophical Studies* 123:47–70, doi:10.1007/s11098-004-5218-x.

—. 2007. "The Bayesian and the Dogmatist." *Proceedings of the Aristotelian Society* 107:169–185, doi:10.1111/j.1467-9264.2007.00217.x.

Williamson, Timothy. 1994. *Vagueness*. Routledge.

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

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