Abstract.
Rationality's Demands on Belief
Alexander Liam Worsnip
2015

It is common in the literature on practical rationality to distinguish between "structural" or "coherence" *requirements of rationality* on one hand, and substantive *reasons* on the other. When it comes to epistemology, however, this distinction is often overlooked, and its significance has not yet been fully appreciated. My dissertation is the first systematic attempt to remedy this. I show that coherence requirements have a distinctive and important role to play in epistemology, one that is not reducible to responding to evidence (reasons), and that must be theorized in its own right. Most radically, I argue that there is a core notion of rational belief, worth caring about and theorizing, that involves only conformity to coherence requirements, and on which – contrary to what most epistemologists either explicitly or implicitly assume – rationality does not require responding to one's evidence. Though this can strike observers as a merely terminological stance, I show that this reaction is too quick, by arguing that there are cases of conflict between coherence requirements and responding to reasons – that is, where one cannot both satisfy the coherence requirements and respond to all of one's reasons. As such, there is no single "umbrella" notion of rationality that covers both conformity to coherence requirements and responding to reasons. As well as trying to illuminate the nature and form of coherence requirements, I show that coherence requirements are of real epistemological significance, helping us to make progress in extant debates about higher-order evidence, doxastic voluntarism, deductive constraints, and (more briefly) conditionalization and peer disagreement. I argue that understanding the relevant norms in these areas as coherence requirements, rather than as claims about reasons, can help to resolve a number of problematic confusions and impasses in these debates.

Rationality's Demands on Belief

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Alexander Liam Worsnip

Dissertation Directors: Keith DeRose and Tamar Gendler

May 2015

Summary Table of Contents

Detailed Table of Contents

Acknowledgments

This dissertation was written with the help of many people.

First and foremost, I must thank my dissertation committee: my co-advisors, Keith DeRose and Tamar Gendler, and my other committee members, Steve Darwall, Bruno Whittle, and (as an external committee member) Ralph Wedgwood. I was lucky to, perhaps unusually, have a committee where every single member, in different ways, was generous with their time, and ultimately made a big contribution to the project and to my intellectual development more broadly. The final dissertation reflects the philosophical diversity of this group, which collectively spans a wide range of philosophical subfields and methodological approaches, and is, I think, all the better for that. Moreover, every member of the committee supported me in other important ways, offering crucial advice, encouragement and big-picture guidance as I shaped the project and, more recently, braved the academic job market. I also learnt a huge amount about teaching both from Keith (as his TA for Epistemology in Spring 2012) and from Steve (as his co-instructor for Rationality in Fall 2013). Both teaching experiences were accompanied by highly stimulating weekly discussions of the material we were covering, and much else besides. And both experiences had a profound effect on this dissertation, helping me to focus the project and to master crucial literature in a deep way by learning how to communicate it to undergraduates.

One of the best things about pursuing my PhD at Yale was the opportunity to get wonderful feedback on my work from many other faculty members who were not formally on my committee. I must single out Dan Greco, Josh Knobe, and Zoltán Szabó in particular for going well beyond the call of duty in providing invaluable help with my work during the time I was at Yale. I also learnt a great deal from classes, discussions and exchanges with a number of other current or former Yale faculty members, including Facundo Alonso, George

Introduction

0.1   A chasm

There is a chasm in the existing philosophical literature on rationality. Specifically, it is a chasm between the literature on doxastic rationality (that is, the rationality of beliefs and other doxastic attitudes) on one hand, and the literature on practical rationality (that is, the rationality of actions and intentions) on the other. Let me explain.

In the epistemological literature on doxastic rationality, the dominant assumption is that rationality consists in responding correctly to your evidence.[1] This *evidentialist* view,[2] in a little more detail, is roughly the following: with respect to any proposition, an agent has at some particular time a body of evidence. This evidence prescribes some particular doxastic attitude toward the proposition in question: a credence; or an "outright" doxastic attitude such as believing, disbelieving, or suspending judgment. And rationality requires the agent to believe what the evidence supports. Indeed, on the evidentialist view, believing what the evidence supports also suffices for believing rationally. In sum, believing rationally is about responding correctly to the evidence you have. Many view this idea, suitably understood, as a truism.[3] In many debates in epistemology, it is simply taken for granted.

Usually, this assumption is tied to – or, indeed, taken to be equivalent to – the idea that evidence (and perhaps, only evidence) provides or constitutes a reason for belief. Thus,

---

[1] See, e.g., Feldman (2000); Kelly (2002); Adler (2002b); Hieronymi (2006); Shah (2003, 2006).

[2] I use the term 'evidentialism' stipulatively. It sometimes refers to a thesis about reasons for belief, not rationality. As I use the term, I am concerned with it as a thesis about rationality. As it happens, those who have most prominently defended evidentialism about reasons for belief have tended also to accept it as a thesis about rationality. But the latter only actually follows from the former if we assume that rationality is a matter of responding to reasons.

[3] Thomas Kelly (2006: §2) calls it "bordering on the platitudinous." Stewart Cohen (forthcoming: 5) writes, "it is hard to see how one could divorce the rationality of belief from considerations of evidence."

the more general assumption behind the assumption that rational belief is a matter of responding to one's evidence is that rationality is a matter of responding successfully to one's reasons. We can call this the "reasons-responsiveness" view of rationality.

Interestingly enough, however, in the literature on practical rationality, the reasons-responsiveness view is a matter of great controversy.[4] Against this model, there is an alternative picture. This alternative picture is endorsed by many philosophers, but has been advanced especially forcefully by John Broome.[5] On the alternative picture, the core, fundamental requirements of rationality are coherence requirements. A coherence requirement, as I am using the term, requires of you that some relation holds between your different mental attitudes. Coherence requirements forbid certain combinations of mental states that are intuitively incoherent in that they don't *fit together* properly. Examples include requirements forbidding contradictory beliefs or intentions, the requirement to take the means to one's ends, and the "enkratic" requirement, which requires one to intend to do what one believes one ought to do. Another name for coherence requirements is "structural requirements". I will use these terms interchangeably. Let us call a view on which *all* requirements of rationality (subject to a minor caveat[6]) are coherence requirements a "coherentist" view.

---

[4] For criticism, see especially Broome (2013: chs. 5-6). See also, e.g., Darwall (1983); Scanlon (1998: ch. 1; 2007); McDowell (1998: ch. 5); Wallace (2001); Setiya (2004, 2007); Kolodny (2005); Mi. Smith (2007); Way (2010); Hussain (ms.: esp. 12-13).

[5] Broome (2013). See also, e.g., Scanlon (1998, 2007); Davidson (2004: ch. 12); Way (2010); Ridge (2014: ch. 8).

[6] The caveat is that there may be some coherence requirements that forbid individual attitudes, if that individual attitude is on its own incoherent. The most obvious example – offered by Broome (2013: 153) – is a rational requirement not to believe a contradictory proposition of the form *p and not-p* (as distinct from having two contradictory beliefs in the two propositions *p* and *not-p*). He suggests that this may be the only coherence requirement that forbids an individual attitude.

This gives rise to the chasm that I referred to at the start. The literature in epistemology on doxastic rationality largely takes the reasons-responsiveness view for granted.[7] The literature in ethics and metaethics on practical rationality at least seriously questions this assumption. Even those that do defend something like the reasons-responsiveness account issue crucial and significant qualifications to it that find no analogue in the epistemological literature on rationality.[8] And in the practical rationality literature, the coherentist view is a very serious contender, perhaps even the majority view.

Some theories even seem to make these divergent assumptions about the doxastic and practical cases in the same breath. On many popular theories of rational decision, there are two crucial inputs: some kind of value function, and some probability function, with rational actors choosing the acts that maximize expected value given these two inputs. At least some philosophers seem to assume simultaneously (i) that the relevant value-function is purely subjective – i.e. that it is given by the agent's actual set of preferences or ends (so long, perhaps as this set is internally coherent), not the set of preferences that would be maximally reasonable for the agent to have – yet (ii) that the relevant set of probabilities is not just the agent's actual subjective credences (even if such credences are internally coherent), but rather the credences that it would be maximally reasonable for the agent to have.[9] In doing so, this view seems to simultaneously encode something like a reasons-responsiveness view about the rationality of credences and a coherentist view about the rationality of preferences.

---

[7] Even John Broome, who offers the most prominent recent defense of the claim that rationality does not require one to respond to one's reasons (Broome 2013: chs. 5-6), focuses overwhelmingly on practical cases in defending this claim. When it comes to the doxastic case, he says that he assumes there are rational requirements connecting your beliefs and your evidence (*ibid*.: 140).

[8] They say that rationality consists in responding to *beliefs* about reasons, or that it consists in responding to some privileged *subset* of one's reasons. See, e.g., Schroeder (2009); Raz (2011: ch. 5); Parfit (2011: ch. 5); Lord (2014b). I address these views in chapter 2.

[9] Cf., e.g., Fantl & McGrath (2009).

This divergence is at least *prima facie* puzzling, and troubling. Many of us would hope that there would be at least some continuity between an account of doxastic rationality and an account of practical rationality. If the reasons-responsiveness view can seem costless in the doxastic case, why does it run into so many objections in the practical case? If the coherentist view seems like a plausible contender in the practical case, why doesn't it even figure in the epistemological literature? In sum: what's going on here? It's this question that motivates this dissertation.

0.2   *Disambiguation?*

Let's begin with a sketch of what I take to be the most deflationary answer to this question. On this answer, the word 'rationality' can simply refer *either* to reasons-responsiveness, or to coherence, and perhaps also to some disjunctive combination of the two.[10] On this view, it is just a sociological quirk of the way that different subdisciplines of philosophy have evolved that epistemologists tend to use the term to refer to reasons-responsiveness, while theorists of practical rationality tend to use it to refer to coherence – and that is what gives rise to the chasm. There is, then, no interesting debate to be had *between* the reasons-responsiveness view and the coherentist view. Call this the *disambiguating response.*

There are some costs to the disambiguating response. On this view, the existing debate in the practical rationality literature between (qualified versions of) the reasons-responsiveness view and the coherentist view will turn out to be a merely verbal one. So, when the interlocutors in that debate take themselves to have substantive arguments for their views, this

---

[10] Cf., e.g., Buchak (2014: 287); Fogal (ms.), Greco (2014a: 201-2, n. 1), and arguably Kolodny (2005: see esp. p. 510). Note that Fogal does not think that coherence is to be understood in terms of coherence *requirements*. Unfortunately, I have not had space to engage with this issue in this dissertation. I take it for granted that coherence can be captured by talk of requirements, but hope to defend this assumption against Fogal's arguments elsewhere.

response diagnoses them as laboring under an illusion. Moreover, it is not obvious that the disambiguating response does a satisfying job of explaining the chasm. For it seems that at least to some degree, the apparently differential uses of the term 'rationality' seem to be reflected not just in academic philosophy but in ordinary talk. Even in ordinary talk, it seems more natural to describe someone who fails to respond appropriately to her epistemic reasons – who fails to believe what her evidence supports – as 'irrational' than it is to describe someone who fails to respond appropriately to her practical reasons (and, in particular, her moral reasons) as 'irrational'. If the chasm is reflected in ordinary talk, it is less plausible that it evolved as a result of the sociological quirks of academic philosophy. So I think that it is at least worth investigating whether there are actually substantive philosophical presuppositions behind the chasm, and if so, whether those presuppositions can be vindicated. I will return to this in chapter 2.

For now, however, let us set these questions aside, and suppose for a moment that the disambiguating response is correct. Even if it is, I want to suggest, this is very far from showing that there are not issues of substantive philosophical interest about the *relationship* between coherence and reasons-responsiveness. And these questions, I want to suggest, have been especially under-theorized in epistemology.

As I said in section 0.1, epistemologists have assumed that doxastic rationality consists in responding to one's evidence. But epistemologists also acknowledge, even if sometimes obliquely, that there are combinations of attitudes that jointly irrational. It is irrational to simultaneously believe some proposition p and to believe its negation. It is irrational to simultaneously believe p while assigning p some vanishingly small credence. It is irrational to simultaneously believe p whilst oneself judging one's reasons for believing p to be no good. So what is the relationship between evidence-responsiveness and these demands of

coherence? At least until very recently indeed,[11] surprisingly little work in epistemology is explicitly devoted to answering this question – indeed, many do not clearly distinguish coherence requirements from evidence-responsiveness in the first instance. But various implicit answers have been given.

One answer effectively tries to reduce evidence-responsiveness to a kind of coherence. This is, I take it, more or less the idea behind the project of the so-called "coherence theory of justification" (see section 0.4 below)[12] and well as that of some forms of subjective Bayesianism (see the conclusion, section a).[13] On this view, for a belief to be supported by the evidence just is for it to cohere in the right way with one's other beliefs (and other doxastic attitudes). So the requirement to believe what one's evidence supports and the various requirements to be coherent ultimately come to the same thing.

A different kind of implicit answer (which I judge to be more popular than the first, at least in the most recent literature) takes it that so-called "requirements" of coherence are ultimately reducible to facts about what the evidence supports. The idea here is that it is a feature of the correct theory of evidential support that one's evidence can never support attitudes that are jointly incoherent.[14] This is easiest to illustrate with respect to the requirement of non-contradiction. Whatever some proposition p is, and whatever one's evidence is, it seems hard to see how one's (total) evidence could simultaneously support believing p, and also support believing not-p. So, if rationality is a matter of taking the attitudes that one's

---

[11] Cf. Pryor (2004: 363-5); Kolodny (2005, 2007), Fogal (ms.), Lord (2014b), Sylvan (ms.), and Easwaran & Fitelson (2015).

[12] Cf., e.g., BonJour (1985).

[13] Cf., e.g., Titelbaum (2010).

[14] Something like this strategy is appealed to by Kolodny (2005) and Lord (2014b). It also seems to be at work in Huemer (2011: 7; compare also p. 3 with p. 4, n, 7).

evidence supports, it will just fall out as a consequence (so the argument goes) that it is never rationally permissible to be in a state such that one has attitudes that violate certain coherence constraints. On this view, there are no *sui generis* coherence requirements, or (to be slightly more metaphysically non-committal) if there are, they are normatively superfluous, since they do not require anything that evidence-responsiveness didn't require already. For every would-be coherence requirement of the form *don't simultaneously have attitude A and have attitude B*, there is a corresponding metaphysical constraint on what combinations of facts about evidential support are possible, of the form *one's evidence cannot decisively support both attitude A and attitude B*, and it won't do any harm to move between these two claims freely.

The two views just sketched clearly differ in a number of crucial respects, but both share the feature of ultimately trying to deny the thought that there are really two different, important normative phenomena here to be pulled apart. On the first view, evidence-responsiveness and coherence ultimately come to the same thing, so there simply is nothing to pull apart. On the second view, evidence-responsiveness entails coherence, but not vice-versa.

The view I will be defending in this dissertation, by contrast, makes the separation between coherence and evidence-responsiveness more radical than these views do. On the view I favor, coherence requirements and substantive reasons are normative phenomena of two fundamentally different kinds, and neither is reducible to the other. Crucial to my argument for this view is my contention, argued for in chapter 1, that coherence and evidence-responsiveness can in fact come into *conflict*.[15] That means that neither coherence nor evidence-

---

[15] This is a revisionary view even relative to those few who do distinguish coherence requirements from evidence-responsiveness, and acknowledge the need to theorize the former. So Easwaran & Fitelson (2015), who nicely distinguish substantive "norms" of evidence-responsiveness from coherence "requirements", assume that it is a decisive objection to a coherence requirement if satisfying it would require one to violate those substantive norms – thus taking it for granted that conflicts between the two are impossible.

responsiveness will guarantee the other. This view makes the separation between coherence and reasons-responsiveness radical even relative to many established views in the practical rationality literature.

Lurking behind the disambiguating response, then, are a number of importantly different views about the relationship between the phenomena being disambiguated. On a view on which coherence is something strictly weaker than reasons-responsiveness, the disambiguation amounts at best to distinguishing a "more demanding" sense of the term 'rationality' – on which it requires evidence-responsiveness *and* (as an automatic consequence) coherence – from a "less demanding" sense of the term – on which it requires coherence only. But if one thinks of things this way, it may not be immediately apparent what the real interest of the second notion would be. Even if there is an important intuition that rationality requires coherence between attitudes, on this view, the first notion of rationality can capture that intuition, whilst doing more besides. There is, on this view, no obvious reason why we need to make room for a notion of rationality as coherence alone.[16]

By contrast, if coherence requirements and reasons-responsiveness can come into conflict, and the latter does not guarantee the satisfaction of the former, things are different. A theory of doxastic normativity that tries to do everything in terms of evidence (that is, in terms of the relevant kind of *reasons*) will then leave something important out. Much of this dissertation will be spent trying to show that what would be left out is a crucial part of the landscape of doxastic normativity – a part that ought to figure in many extant epistemological debates – and that the persistent tendency of epistemologists to run evidence-responsiveness

---

[16] So, for example, Kurt Sylvan (ms.: 5) writes, "many meta-ethicists understand rationality very narrowly. While they hold that rationality requires responding to apparent reasons, many will… view this requirement as a coherence requirement. Epistemologists are unlikely to find this interesting. Coherentism about anything other than coherence is now widely rejected."

and coherence requirements together, or to try to translate talk of one directly into talk of the other, has a distorting effect on many of these debates. Moreover, if one tries to construct a kind of "umbrella" notion of 'rationality' that covers *both* reasons-responsiveness and coherence, it will be one that (given the possibility of conflicts between the two) sometimes issues demands that are not jointly satisfiable.

Where does this leave us with the question of how to use the general term 'rationality'? In some ways, I agree with the proponent of the disambiguating response: there is clearly a strand of philosophical (and ordinary) talk that uses 'rationality' to refer to evidence-responsiveness in the context of belief. And I certainly don't want to deny that evidence-responsiveness is an important part of a complete theory of epistemic normativity. However, that is all compatible with making the revisionary (at least relative to epistemology) *proposal* that 'rationality' be used to refer to the satisfaction of coherence requirements only, with the vocabulary of reasons, rather than rationality, being used to capture the normative importance of evidence-responsiveness. This usage would, in my view, be clearer, and in many respects prevent us from falling into philosophical mistakes that are themselves substantive and not merely verbal. In choosing terminology, it is reasonable to consider whether some possible set of terminology saves us from running distinct things together, or from lacking a word for something distinctive and important.

To be clear, I do not intend to stipulate this proposal about the use of 'rational'. Rather, I intend to *argue* for it by showing the importance of coherence requirements, the explanatory and philosophical work that a notion of rationality as pure coherence can do, the distinctness of claims about coherence from claims about evidence-responsiveness, and the possibility of conflicts between the two. The case for this proposal, then, will emerge gradually throughout chapters 1 and 2, rather than being a terminological stipulation from the outset.

19

In the sense just explained, then, one can think of this dissertation as arguing for a coherentist view of doxastic rationality. But even if one rejects my proposal about restricting 'rationality' to the satisfaction of coherence requirements alone, that doesn't obviate the underlying point about the distinctness of coherence requirements and evidence-responsiveness, and the importance of pulling them apart and giving an adequate account of the former and the role they play in epistemology. These too are central aims of this dissertation.

0.3 Substantive reasons versus coherence requirements – a more detailed first pass at the distinction

It's now time to say a little bit more about the distinction between substantive reasons and coherence requirements, and how to characterize the latter in particular.

I will assume that reasons for doxastic attitudes typically involve a two-place relation: some proposition (or fact/appearance/seeming/intuition/attitude) speaks in favor of some attitude. Claims about reasons are thus typically "narrow-scope", in the sense that they take the following generalized form, where p and q are propositions:

p is part of your evidence → you have a reason to take attitude A toward q

The claim is narrow-scope because the relevant normative notion – that of your having a reason – takes scope only over the attitude mentioned in the consequent of the conditional; it does not take scope over the whole conditional. Several other related points are salient. First, what you have reason to do here is to take some particular, determinate attitude A toward q – not to make true some conditional that reflects some relationship between multiple attitudes. Second, your having reason to take this attitude depends on what your evidence is. The above

conditional says that if p is part of your evidence, then you have a reason to take attitude A toward q. But you might not have any reason to take attitude A toward q if p is not part of your evidence. What evidence you have, in turn, is a matter of how the world is and how you have encountered it. So saying whether you actually do have reason to take attitude A toward q will often depend on our knowing what your evidence is like, which is a (partly) empirical and contingent matter.

By contrast, coherence requirements specify which attitudes (or lack thereof) it is rational or irrational to *combine*. The fundamental coherence requirements, in my view, are typically "wide-scope", in the sense that they take the following generalized form:

Rationality requires (you have attitude A toward p → you have attitude B toward q)

The requirement is wide-scope because the relevant normative notion – that of a rational requirement – takes scope over the whole conditional, not over the attitude in the consequent of the conditional. Correspondingly, what you are rationally required to do here is to make true the material conditional within the scope of the requirement. The idea is *not*, then, that when you have attitude A, there is an unconditional requirement to take attitude B – as if attitude A has some fixed authority or unrevisability. Rather, you can just as well make the material conditional true by giving up attitude A. As a result, one cannot "detach" a requirement to have attitude B merely from the fact that one does have attitude A. If the conditional form of the requirement gives it a misleading impression of non-symmetry here, it may help to reformulate the material conditional as the equivalent disjunction:

Rationality requires (you have attitude B toward q ∨ you do not have attitude A toward p)

It can also help to reformulate the requirement as one to avoid the relevant attitudinal mismatch:

Rationality requires that it's not the case that (you have attitude A toward p ∧ you do not have attitude B toward q)

Again, several points bear noting. What you are rationally required to do when you are under a coherence requirement is to make true (or false) some proposition that is a logical combination of potential attitudes you might have. More colloquially, you are required to ensure that the right relationships between your attitudes (and lack thereof) hold; hence the label 'coherence' or 'structural'. Individual coherence requirements do not speak in favor of any one determinate attitude. To be clearer here, let us work with a specific putative coherence requirement instead of with the general illustration of the form of a coherence requirement above. We will work with (a rough formulation of) the instrumental requirement:

(IR) Rationality requires of you that ((you intend end E ∧ you believe that M-ing is a necessary means to E-ing) → you intend means M).

This requirement is entirely silent on whether you should satisfy it by taking means M, or by giving up end E, or indeed by giving up your belief that M-ing is a necessary means to E-ing. Other normative considerations may settle that, but the coherence requirement itself does not.

It thus does not speak determinately in favor of any one of these individual attitudes. Relatedly, in contrast to your having a reason to believe something, your being subject to this coherence requirement is not contingent on empirical matters about how the world is and how you have encountered it; it is not even contingent on what end E and means M are! Whatever the end, whatever the means, whatever their results and what you know about their results, you are incoherent (and so irrational) if you intend the end, believe the means is necessary for the end, but fail to intend the means.

In what sense are coherence requirements non-substantive, as the contrast with substantive reasons implies? I do not mean to imply that they are *insubstantial*, or normatively unimportant, or somehow trivial. I just mean to indicate that coherence requirements are silent on the substantive merits of particular individual attitudes. The instrumental requirement is entirely silent on whether end E or means M are worthwhile, for instance. Substantive reasons for pursuing such ends and means are, obviously, not silent on this matter.

For all I have said here in this introduction, it remains a matter for debate whether the instrumental requirement (and other requirements like it) are correctly formulated as wide-scope coherence requirements, or whether they should instead be formulated with narrow-scope.[17] I return to the issue of scope in much greater detail in chapter 3. However, for now I will assume that the correct way to formulate these fundamental requirements – in contrast to claims about reasons – is indeed with wide-scope. Although chapter 3 will complicate my stance, the position I argue for leaves this fundamental claim intact.

---

[17] For philosophers who endorse the wide-scope view, see Broome (1999, 2007, 2013: ch. 8); Bratman (2009); Brunero (2010, 2012); Dancy (1977, 2000: ch. 3); Darwall (1983); Gensler (1985); Greenspan (1975); Hussain (ms.); Rippon (2011); Setiya (2007); Scanlon (2007); Mi. Smith (2007); Wallace (2001); Way (2010, 2011). For dissenting defenses of narrow-scope views, see Schroeder (2004), Kolodny (2005) and Lord (2014a).

However, I do briefly want to illustrate (in a way that will be, of necessity, sketchy) how distinguishing (wide-scope) coherence requirements from substantive reasons has had payoffs in the literature about practical rationality of the last few decades, to give a taste of what I hope to do for epistemology in this dissertation.

Intuitively, there is something wrong with an agent who fails to act on her preferences, and to take the means to her ends. If substantive reasons are all one has in one's normative ontology, it seems like accounting for this requires one to say that her (actual) preferences and ends give her reasons – perhaps, if one is inclined towards parsimonious accounts of reasons, that they are the *only* things that give her reasons. Yet this stance appears to commit one to a rather subjective view of reasons. On this view, my merely having some preference or end can give me a reason to act on that preference or end – no matter how crazy, cruel, immoral, or worthless the end. But equally, the thought that one's actual preferences or ends are just irrelevant to one's normative situation also seems wrong. Something does seem to go wrong when one fails to align one's intended ends and intended means in the right way, for instance.

We can resolve this apparent impasse by saying that the relevant normative phenomenon here is not a substantive reason but rather a coherence requirement. When one intends E, and believes that M-ing is a necessary means to E, but fails to intend M, one violates a coherence requirement. That is so even if E-ing is crazy, cruel, immoral or worthless. And it is so even if both M-ing and E-ing are optional, so far as one's substantive reasons go – such that both intending E and failing to intend M are *individually* reasonable. Once we allow coherence requirements as a phenomenon distinct from substantive reasons, we can say all of this without saying that one's intentions or preferences just *give* one reasons.[18]

---

[18] This fundamental point is made nicely by Darwall (1983: ch. 4), amongst others.

Similarly, the distinction between coherence requirements (and a notion of rationality as satisfaction of coherence requirements) and substantive reasons enables us to reach a nice compromise on what to say about agents whose mental attitudes *do* intuitively fit together in the right way, but who are inadequately responsive to moral considerations. Take, for example, someone who is inadequately kind in his dealings with others. If we are determined to tie rationality to responsiveness to one's reasons, then it looks like we must say one of two things. On one hand, we could say, with so-called "moral rationalists", that this individual's lack of kindness makes him *irrational* – a claim that many have wanted to resist. Or we could say, with so-called "reasons internalists", that in fact this individual has no reason to be kind (unless, perhaps, he has some kind of deep desire to be kind, or a deep desire that cannot be satisfied unless he is kind – and he might well lack such a desire).

However, if we separate rationality (as coherence) from reasons, we can again arrive at a highly reasonable compromise. The individual in question does fail to respond to reasons that are in fact genuine and present, his desires notwithstanding. But he is not necessarily irrational (at least not in a core and important sense of 'irrational') as a result of this.[19] So again, we can steer a course between two extremes to a reasonable middle ground, avoiding a false dichotomy and making substantive progress in an important philosophical debate.

As I'll be arguing in this dissertation, I think there are parallel situations in epistemology where a failure to distinguish reasons and coherence requirements has hampered progress. For example, I'll suggest that something like this is true in debates about higher-

---

[19] Early reasons internalists were keen to preserve the thought that is not necessarily irrational to be immoral. See Williams (1981): 110; 112; Foot (1972: 161-2). So perhaps admitting this thought gives them at least some of what they originally wanted, or even what they meant to convey when they said that morality does not necessarily give us reasons. In my view, this latter claim over-reaches, but the kernel of insight in Williams and Foot is that morality is not necessarily *rationally required* of us. For similar points about reasons internalism, see Scanlon (1998: 27); McDowell (1998: ch. 5, esp. 110); Setiya (2004: 269); Ridge (2014: 230-1, 233-4). See also section 2.4 below for further discussion.

order evidence (chapter 1), in debates about doxastic voluntarism (chapter 4), and (more tentatively) in debates about belief updating of various kinds (conclusion, sections a and b). So there are real payoffs, I'll argue, of distinguishing coherence requirements and substantive reasons in epistemology.

At a more general level, a tendency to try to work with only one normative notion can obscure matters in epistemology just as it can in ethics. For example, Roger White (2007: 126), targeting something like a coherentist view of doxastic rationality, complains of coherentism that you and I might have "radically different points of view," yet both be ideally coherent. He concludes that that coherentists must say that "there is no epistemic criticism that either of us can offer the other." But that assumes that all epistemic criticism amounts to accusations of irrationality, and that there are not other normative notions that can do important work. This is just the analogue of complaining, in the practical case, that someone who does not convict the immoral agent of irrationality cannot offer any normative criticism whatsoever. Similarly, White says that if I think of two sets of beliefs as equally rational, one "might as well flip a coin to decide which to adopt" (*ibid*.: 127). Again, this assumes that rationality is the only property worth having for one's beliefs.

0.4   *What the coherentist view is not*

I will have more to say about some distinctive features of coherence requirements – and what makes them different from substantive reasons – as the argument of the dissertation develops; in particular, in chapters 2 and 4. In the meantime, I hope I have said enough about the distinction to work with for now. I will now issue several clarifications about what coherence – and the coherentist view of rationality – are not.

a. Not "subjective" reasons

First, I want to be clear that the distinction between coherence requirements and reasons is not equivalent to the distinction between so-called "subjective" and "objective" reasons. There are multiple different ways of cashing out the distinction between objective and subjective reasons.[20] What the different ways of drawing the distinction have in common in that whereas objective reasons are supposed to track what one has reason to do given some more perspective-independent body of information, subjective reasons are supposed to track what one has reason to do given some more perspective-dependent body of information. For example, objective reasons might be the reasons one has given all the facts, and subjective reasons might be the reasons one has given one's beliefs.

One might think that those who, like me, distinguish rationality (as coherence) and responding to reasons are effectively proposing that *rationality* is to be associated with responding to subjective reasons, where these are extremely perspective-dependent, whereas *reasons proper* are to be understood as maximally objective, perspective-independent reasons.[21]

While some philosophers hold something not unlike this view,[22] let me be clear that it is *not* the view I am defending here. In separating rationality and reasons, I am not restricting the term 'reasons' to maximally objective, perspective-independent reasons. Indeed, I will

---

[20] For a nice overview, with quotations, of how different philosophers draw the distinction in radically non-equivalent ways, see Feldman (1988). Indeed, the fact that there are so many non-equivalent ways of drawing the distinction should make us suspicious that any binary distinction will exhaust the important notions here. For this reason, amongst others, I favor a contextualist semantics for normative terms that acknowledges this broader range and tries to unify the many usages under a single semantics, rather than positing a binary ambiguity. See section 2.6, part c, and chapter 3, below.

[21] This seems to be how Gibbons (2010) interprets the proposal to distinguish rationality and reasons. He shows without much difficulty that there are various problems with this proposal. But it seems to me that his target is a straw man of the view that rationality and reasons come apart.

[22] See section 2.6, part b, below.

argue in section 2.6 part b that even reasons-responsiveness accounts of rationality which use more "subjective" notions of 'reason' fail. Rather, the notion of a coherence requirement is to be distinguished from the notion of a reason quite generally. As defenders of coherence requirements have stressed since Broome (1999), and as I tried to make clear in the last section, what distinguishes a coherence requirement from a reason is not its perspective-dependence *per se*, but rather its wide-scope logical structure. A coherence requirement merely bans certain impermissible combinations or "mismatches" of attitudes (or, in some cases, absences of attitudes), and does not speak determinately in favor of some individual attitude. By contrast, even in the case of *subjective* reasons, it is still true that if and when an attitude A provides a (subjective) reason for another attitude B, one can correctly respond to this reason only by taking up attitude B. Giving up attitude A would not be a way of responding to the reason that attitude A gives one to have attitude B – even if one has some other reason to give up attitude A.[23] So subjective reasons do not have the same logical structure that coherence requirements do.

b. *Not (logical) consistency*

The notion of coherence here is also not merely that of consistency. As I have said, coherence requirements are requirements that take scope over some combination of attitudes, specifying when it is rational or irrational to hold such attitudes jointly. Broadly speaking, they ban

---

[23] This explanation of the difference between subjective reasons and coherence requirements leaves open the possibility that one notion might be able to make the other redundant, by accounting for the data it was supposed to account for. One might claim, for example, that mere beliefs do not in any good sense "give one" reasons, subjective or otherwise, but merely feature as part of coherence requirements, and that the former idea is just a confused way of getting at the latter. This seems to be close to Jonathan Dancy's view (see Dancy 1977, and especially Dancy 2000: ch. 3).

combinations of attitudes (or lack thereof) that do not fit together properly. But the reason that the attitudes do not fit together properly need not be traceable to any logical inconsistency in their contents. We have already seen an example of a coherence requirement – the instrumental requirement – which one can fail to obey without any logical inconsistency in one's attitudes. In fact, in chapter 5, I will argue that deductive consistency is *not* actually a requirement of coherence. So deductive consistency is certainly not sufficient, and I will argue not even necessary, for coherence.

c.   *Not the "coherence theory of justification"*

When epistemologists hear the term 'coherentism', they will likely think of a view that is not at all new – the so-called "coherence theory of justification."[24] It is crucial to see from the outset how this is not the view that I am defending. The coherence theory of justification, as traditionally understood, is the view that the epistemic justification of a belief can only be understood against the background of one's total set of beliefs, and that the more coherent one's total set of beliefs are, the more justified they are. Here, coherence is understood as a property that identifies an *evidential support relation* between propositions.[25]

Ultimately, then, the "coherence theory of justification" is yet another view that attempts to keep coherence and evidence-responsiveness together, rather than pulling them apart. It does so by trying to understand evidential support in terms of a kind of coherence. This is not my view. I make no commitment to understanding evidential support in terms of

---

[24] See, e.g., BonJour (1985).

[25] Cf., e.g., Harman (1999: 33-35).

coherence. If 'epistemic justification' refers to the property of being supported by evidence, then I make no commitment of understanding justification in terms of coherence either. (See section 0.7 below for more on 'justification'.)

Moreover, I do not think of coherence as a gradable property of a set of beliefs that is about how well they positively "support" each other. Rather, I aim to enumerate precise requirements that state individual requirements of coherence, each of which is stringent. As a set these requirements are a diverse bunch. While we can define a single property of coherence in terms of the satisfaction of this full set of requirements, my view is that it is easier to think about coherence requirements by considering them individually than by beginning with some loose, undifferentiated notion of coherence. As I use the term, two attitudes are only incoherent when there is a requirement that forbids holding them jointly. The mere sense that they do not sit together perfectly is not enough.

In this dissertation, 'coherentism' refers to the view that rationality is about the satisfaction of coherence requirements – and not to the traditional "coherence theory of justification".

d. Not (merely) probabilistic coherence

There is yet another way in which the term 'coherence' is sometimes used in epistemology (particularly formal epistemology): to refer to the *probabilistic* coherence of degrees of belief (credences). Degrees of belief are coherent, in this sense, if they conform to the probability axioms. So, for example, it is probabilistically incoherent (in this sense) to have credence 0.9 in some proposition p, credence 0.9 in some other proposition q, but credence 0.95 in the conjunctive proposition (p ∧ q).

30

I take it to be a substantive question whether there is a genuine rational requirement that one's degrees of belief be probabilistically coherent.[26] I do not take a stand on this debate in this dissertation. The putative requirement of probabilistic coherence is a *putative* coherence requirement, in my sense of the term, since it specifies a particular combination of mental states that (by the lights of the requirement) it is irrational to hold jointly (namely, any two or more credences that fail to conform to the probability axioms). If it is a genuine rational requirement, then it is a genuine coherence requirement. So, the view that it is rationally required that one's degrees of belief be probabilistically coherent (so-called "probabilism") is consistent with the view I defend, that all rational requirements are coherence requirements. But so is the view that probabilism is false. In this dissertation, I only call combinations of mental states 'incoherent' when there is a *genuine* rational requirement (of coherence) that forbids holding them jointly.[27] So, in my sense of 'incoherent', it is a substantive question whether (all) instances of probabilistic "incoherence" are *genuinely* incoherent. Being a coherentist about rationality, in my sense, leaves one free to either accept or deny probabilism.

What I do commit myself to is the claim that, if probabilistic coherence is a rational requirement (of coherence), it is not the *only* rational requirement (of coherence). Nor is it the only rational requirement governing doxastic states. I will be defending numerous other rational requirements in this dissertation, all of which (if they are, as I will argue, genuine rational requirements) are coherence requirements. So at the very least, I am committed to the claim that (doxastic) coherence is not *exhausted* by probabilistic coherence.

---

[26] For relevant debate, see e.g., Foley (1993: sec. 4.6); Joyce (1998); Christensen (2004: chs. 5-6; 2007) and Titelbaum (2015b). One major alleged drawback of the view that it is rationally required to be probabilistically coherent is that it commits us to saying that it is rationally required to have credence one in all logical truths.

[27] Contrast Foley (1993: sec. 4.6), who argues that one can be "rationally incoherent", by which he means, probabilistically incoherent but rational. Translated into my terminology, Foley does not think that probabilistic coherence is a *genuine* rational requirement.

0.5 Domains of rationality and domains of philosophy

Although the issues that I pursue in this dissertation concern rational belief first and foremost, and thus engage with topics and work that are traditionally classified as epistemology and (to a lesser extent) philosophy of mind, I will frequently draw upon analogies with practical rationality and work done in ethics and metaethics (broadly construed). It is my view that the study of theoretical rationality and the study of practical rationality each have much to learn from the other, and that filing them under distinct areas, "epistemology" and "ethics" respectively, is misleading and bad for advancing each field. Fortunately, the tendency to see things this way is in decline. Nevertheless, as the chasm that I drew attention to at the start illustrates, the two domains are still very far from being unified. There is, I think, much yet to be learnt in the theory of rational belief from existing work on practical rationality. I hope that this dissertation makes some progress towards this.

Indeed, there is a sense in which, on a wide-scope coherence account of rationality, the binary distinction between doxastic and practical rationality is itself undermined. After all, remember that on this account, rational requirements forbid certain combinations of attitudes. So by contrast with a narrow-scope account, it's not as if we can neatly sort them into the requirements that mandate beliefs, and those that mandate intentions. For they can mandate or forbid certain *combinations* of beliefs and intentions. Indeed, we have already met with an example of a wide-scope requirement of exactly this kind: the instrumental principle, which relates a *belief* about the necessary means to an end of yours to an *intention* to take that means. And it doesn't stop there: wide-scope requirements can relate beliefs to hopes, to fears, to regrets, to suppositions, and in principle to any propositional attitude whatsoever.

The situation is even worse for accounts that put a big emphasis on a distinction between "epistemic rationality" and something else. The doxastic/practical distinction tracks a difference in the target of evaluation: doxastic rationality concerns beliefs, whereas practical rationality concerns intentions or actions. On the other hand, the epistemic/non-epistemic distinction, as I understand it, tracks a difference in what that target is evaluated in terms of. An epistemic reason for belief might be glossed as one concerning one's evidence for the proposition believed, whereas a non-epistemic reason for belief (if there be any such reasons) may concern something else, such as the pragmatic value of having the belief.

The epistemic/non-epistemic distinction thus makes good sense as a distinction between different kinds of reasons for belief; indeed I will often talk of epistemic reasons.[28] But if we think of rationality as consisting in coherence between attitudes, rather than in responding to reasons, it makes very little sense as a distinction between different kinds of rationality. On the coherentist view, evaluating an agent's rationality is not a matter of evaluating her attitudes in terms any kind of reason, epistemic or otherwise. Rather, it is a matter of seeing whether she fulfils particular coherence requirements. So it doesn't make sense to talk of epistemic rationality as opposed to some other kind of rationality, except insofar as the epistemic/non-epistemic distinction is being conflated with the doxastic/practical distinction. So I will not use the term 'epistemic rationality' in this dissertation.

---

[28] Use of the qualifier 'epistemic' can, however, have a kind of trivializing effect in certain debates. Several writers (e.g. Feldman 2000: 678) maintain that when they claim that all reasons for belief are evidential, they are restricting themselves to epistemic reasons. But if what makes something an epistemic reason is just that it concerns one's evidence, then this claim is a (not-so-)concealed tautology, and consequently of no interest at all. (Thanks to Keith DeRose for making this point to me forcefully; see also Cohen forthcoming.) What is much more interesting is the idea that the very idea of a non-epistemic (non-evidential) reason for belief does not make sense (cf. Shah 2006; Parfit 2011: Appendix A).

0.6   *Evidence, reasons, and 'ought'*

Next, let me say something about how I understand the notion of evidential support. I assume that agents can possess bodies of evidence. I want to try to stay neutral on what kind of items get into this body of evidence – propositions, facts, or even (if you prefer) sensations and appearances. I will often speak as if a body of evidence consists of a body of propositions. But in general, I hope that not too much will turn on this.

As I have already indicated, what the evidence supports can be understood as the attitude or attitudes that your total body of evidence gives you most reason to take. There are several important things to note here. First, the word 'support' can be used in either a *pro tanto* or a *pro toto* sense – we can refer to evidence as providing partial, defeasible *pro tanto* support for a particular doxastic attitude, or to the doxastic attitude that your evidence supports *pro toto* – that is, on balance. As my definition makes clear, I am using the term 'support' in the *pro toto* sense (unless I specifically indicate otherwise).

Second, I am understanding evidence as providing support for *attitudes*, rather than for *propositions*. I do not deny that talk of evidence supporting propositions also makes sense. But for various reasons, I prefer to talk of evidential support for attitudes. It makes things more parallel to the practical case, where it is clearly attitudes (and actions) that are supported by reasons. There is another reason for talking of support for attitudes as well. When it comes to propositions, there are only two candidates for being supported: p and its negation. But when it comes to attitudes towards some proposition p, there are *three* (outright) attitudes that can be supported by one's total evidence: believing p, disbelieving p (believing not-p), and suspending judgment about p. It is important that we don't lose sight of the possibility of one's total evidence supporting suspending judgment with respect to some proposition, or throw that case together with the case where it supports actively disbelieving the proposition.

Talking of support for attitudes (rather than propositions) makes clearer room for the full range of possible doxastic responses, and for support for each of them.

Third, and importantly, the definition makes explicit that it is a *normative* question whether some proposition (or fact, or other entity) evidentially supports an attitude. Just as it is a normative question whether some proposition (or fact, or other entity) gives one reason to perform an action, it is a normative question whether some proposition (or fact, or other entity) gives one reason to take a doxastic attitude. It is true, and important, that a proposition's evidentially supporting a belief can be parsed as its in some way making the truth of the proposition believed more likely (cf. section 1.3 below). But the notion of "likelihood" on one's evidence cannot be understood in terms of some purely descriptive notion such as statistical frequency.[29] The disagreement, say, between dogmatists and their opponents about whether an appearance of a hand evidentially supports believing the proposition that one has hands over believing the proposition that one is a handless brain in a vat is not a descriptive disagreement about the statistical correlation between hand-appearances and real hands. Rather, it is a normative disagreement through and through. Likewise for debates over whether a theory's simplicity is evidence for its truth, over whether the existence of the correct conditions obtaining for life on earth is evidence of God's existence, or over whether the fact that one's upbringing explains one's having some belief provides defeating evidence against that belief. These are normative debates; figuring out evidential support relations is thus, at least in part, a normative enterprise, and an *extremely* tricky one at that.

This fact will be important at several points in this dissertation. Failure to realize (or at least to sufficiently appreciate) it has, I think, led to some profound philosophical mistakes.

---

[29] See also Kelly (2007).

It has led philosophers to resist thinking of core epistemology as a truly normative subfield of philosophy, or to think of epistemic normativity as somehow unproblematic in a way that moral or other kinds of practical normativity are supposed not to be. Some philosophers seem to assume that when it comes to the epistemic case, normative inquiry is exhausted by the identification of some general norm like "believe what your evidence supports", and from there on, figuring out what the evidence supports is a descriptive (and perhaps not very difficult) enterprise. But such a general norm is on its own an empty formality, like "do what you have most practical reason to do." The details of working out your epistemic reasons come in figuring out the evidential support relations – which propositions (or facts) support which doxastic attitudes – just as the details of working out your practical reasons come in figuring out which propositions (or facts) support which actions.

Relatedly, it is possible to be suspicious of whether there really are objective evidential support relations – whether there are mind-independent normative facts specifying which individual propositions count in favor of which particular doxastic attitudes and with what weight. Such suspicion is exactly parallel to suspicion that there are mind-independent normative facts specifying which individual propositions count in favor of which particular actions and with what weight. In other words, just as there is a metaethics of practical reasons, there is an (underexplored) metaethics of epistemic reasons, and in both cases robust realism is not the only option. I think that in general, we should expect reasons to be suspicious or to be unsuspicious in one of these cases to carry over to the other. It is hard to see, for example, why positing an objective, normative reason-giving relation between facts and actions is supposed to be "spooky", or to involve "queer" facts and properties "out there in the world", but positing an objective, normative evidential support relation between facts and doxastic

attitudes is not supposed to involve such pitfalls. Overlooking the normativity of the evidential support relation can obscure the parallel here.

That said, in this dissertation I will be making the procedural assumption – with the bulk of my epistemological opponents – that we can indeed make sense of the idea of objective evidential support relations between propositions and doxastic attitudes. This assumption is not universally shared.[30] But it is dialectically warranted. Suppose that there are *not* in fact objective evidential support relations. Then there are no objective facts about whether agents respond correctly to their evidence to ground a theory of rationality in terms of (objective) evidence-responsiveness. Nor could it then be a mark against the coherentist theory that it fails to count such failures as irrational, since there are no such failures, really. Of course, we could assess whether agents successfully meet their *own* evidential standards. But that would just be to ask whether they are (in one dimension) coherent. If the purported rational requirement to believe what one's evidence supports turned out simply to be a requirement to meet one's *own evidential standards*, then it turns out itself to be a coherence requirement, and so is no threat to the coherentist view of rationality.[31]

I will also assume that there is some at least some value of 'ought' on which what one ought to do or believe is equivalent to what one has most reason to do or believe. This 'ought' is often called the "all-things-considered" 'ought'. On one plausible proposal, reasons are *pro*

---

[30] For example, Titelbaum (2010), along perhaps with some other subjective Bayesians, rejects it (I do not think rejecting it need be a commitment of Bayesians – see the conclusion, section a). Titelbaum holds a subjectivist view of evidential support where what it is for a proposition to support a doxastic attitude is for it to support that attitude *according to the standards of the doxastic agent*. This view, it must be said, is *radically* error-theoretic as regards ordinary judgments (again, see the conclusion, section a). As in the practical case, there are potential alternatives to this radically subjectivist view that fall short of robust realism – expressivism, constructivism, and so on – that may not have this extreme result.

[31] Though I do think there are good reasons to line up one's metaethical story about epistemic and practical reasons, I do not think that one's metaethical story about coherence requirements need be the same as one's metaethical story about reasons. See section 2.1 below.

tanto – in that they defeasibly count in favor of particular attitudes and actions – and you all-things-considered ought to do (or believe) whatever you have most reason to do. As I touch on in chapter 3, I subscribe to the view that 'ought' is semantically context-sensitive – in that it takes a different semantic value depending on conversational context. So not every value of 'ought' is equivalent to 'most reason'. But plausibly, there is at least one value of 'ought' that is.[32]

0.7   The epistemological context of this dissertation

In part 0.1, I said that the claim that doxastic rationality is matter of believing what the evidence supports is hegemonic in epistemology. However, the reader might wonder about some prominent epistemological theories that might at first appear to be exceptions to this trend. I have already considered the so-called "coherence theory of justification" in section 0.4. Here I focus on two other theories: instrumentalism, and reliabilism.

First, there is the so-called *instrumentalist* view.[33] On the instrumentalist picture, rationality is about satisfying your goals. So, rationality requires you to believe something just if having that belief would help to satisfy your goals. However, as it is usually developed, the view still claims that *epistemic* rationality is nevertheless a matter of one's beliefs being

---

[32] Making things more complicated here is the question of whether 'reason' and 'evidence' are themselves context-sensitive. See Henning (2014) for the view that 'reason' is, and Neta (2003) and Wedgwood (2006: 154-5) for the view that 'evidence' is. (There could even be a usage of 'reason' that can just be defined in terms of coherence, as I will argue is the case for 'ought' in chapter 3.) I am assuming that this makes more trouble for my opponents that it does for me. If there are many different possible semantic values of 'evidence' and 'reason', the question arises as to which is supposed to be the relevant one for the proponent of a reasons-responsiveness view of rationality. I assume that she will want to say that there is some central sense of 'evidence' which is the one relevant to rational belief. I'm not sure if this maneuver actually solves all her problems, but I'll grant the assumption for the sake of argument. Thanks to Ralph Wedgwood for discussion of this point.

[33] See, e.g., Foley (1993); Nozick (1993: ch. 3); Kornblith (1993). Apparently 1993 was a great year for instrumentalism!

supported by the evidence. Foley (1993), for example, holds that what is distinctive of epistemic rationality is that it is concerned with goals related to the pursuit of truth, and that believing in accordance with one's evidence is the means to this pursuit.[34] Moreover, even to the extent that non-epistemic goals are let into the picture, the most prominent versions of the instrumentalist view still qualify as reasons-responsiveness views more broadly.[35] On Foley's picture, again, your goals provide you with reasons, and rationality consists in responding to these reasons.

What about reliabilism – often taken so be a rival to so-called "evidentialism"? Reliabilists such as Alvin Goldman (1986) claim that a belief is justified to the extent that it is formed by a reliable belief-forming process – which is supposed, by Goldman and his critics, to be independent of whether it is supported by the evidence. So justification is not a matter of evidential support. However, Goldman explicitly distinguishes rationality and justification.[36] So it does not follow that he breaks with the epistemological consensus of treating *rationality* as being a matter of evidence-responsiveness; in fact, this is just what his distinction seems to be designed to allow.

In fact, it is not obvious that reliabilism even entails that *justification* is not a matter of evidence-responsiveness. This issue is rather closely bound up with the wars between so-called "internalists" and "externalists" in epistemology, which exhibit an interesting historical pattern. Originally, reliabilism was a thesis about knowledge: roughly, that a true belief is

---

[34] For a critique of the instrumentalist view of epistemic rationality, see Kelly (2003). My sense is that the instrumentalist view is presently out of fashion.

[35] Steve Darwall rightly pointed out to me that one can imagine an instrumentalist theory of rational belief that construes the relationship between goals and attitudes in a more coherentist manner, rather than thinking of goals as providing reasons for attitudes.

[36] Goldman (1986, 2009).

knowledge just if it is reliably formed. This is taken to be an "externalist" thesis about knowledge because, in some important sense, this reliability is thought to be something external to the believing agent's own mind. Now, in Goldman's earlier formulations of this view, he takes the view to entail that one can have knowledge without *justification*.[37] The tacit assumption here seems to be that justification is a purely internal matter, so to the extent that one takes an externalist view of knowledge, one divorces knowledge and justification. Later, however, Goldman sees that he can go externalist (and reliabilist) about justification just as he can go externalist about knowledge, thus keeping knowledge and justification together.[38] But even with this established, many assumed that at least *evidence* had to be a purely internal matter. And so, correspondingly, they assumed that an externalist view of justification divorces justification and evidence. "Internalism" and "evidentialism" about justification started to be treated as near-synonyms.[39] Unsurprisingly, however, eventually some saw the natural next step, namely to go externalist about *evidence*.[40] If one goes externalist about evidence, it is not obvious that an externalist, reliabilist view of justification must divorce justification from evidence-responsiveness.

Nevertheless, whether evidence is best construed along externalist lines is a substantive question. So let me make the following proposal, which I hope will shed some light on the more general question of how to fit 'justification' into the terminological framework developed here. Really there are three questions that need to be distinguished (at least intensionally): whether one's beliefs have the property of conforming to coherence requirements, whether

---

[37] See, e.g., Goldman (1976).

[38] See Goldman (1986).

[39] The paradigm exponents of the two as a unified view being Conee & Feldman (1985).

[40] Williamson (2000); Pritchard (2002).

one's beliefs have the property of being *pro toto* supported by the evidence, and whether one's

beliefs have the property (or properties) that are (in addition to truth, and perhaps anti-Gettier

luck) required for them to count as knowledge. I think that calling a belief 'justified' is most

naturally interpreted as involving an affirmative answer to the second question, and when I

use it that will be my meaning.[41] But I can see that it might refer to one of the others. I am

hopeful, however, that it does not refer to some fourth idea, and that the existing tripartite

distinction is enough to map out the possible views.[42]

With that clarification in place, I want to say a little about how the view that I will

defend in this dissertation fits into the historical picture of the internalism-externalism wars

that I sketched a moment ago.[43] One can see that historical picture as a story of philosophers

gradually pushing the limits of externalism further and further: first knowledge, then

justification, then evidence. At each stage, those with internalist leanings are tempted to say,

"well, OK, perhaps we can be externalists about *this*, but surely we must be internalists about

that?" And at each stage the externalist laughs and rolls back the internalist frontier yet further.

I feel conflicted about this dialectic. On one hand, there does seem to be something

that the internalist intuition gets at – that there must be at least *some* normative concept in

---

[41] There is at least one good reason to think that in natural language 'justification' and 'rationality' are different. Justification – like evidence – is something that it is always possible to have more of, in the sense that even if you are responding perfectly to what you have before you, you could have more before you to justify your actions or beliefs. But rationality does not seem to be this way, regardless of one's theory of rationality. In some sense rationality is about making the best of what you have. If you are making the best of what you have, then having more before you might make you more justified, but not more rational.

[42] Whether any of these properties are *extensionally* equivalent is a substantive question. Those that hold the aforementioned "coherence theory of justification" take at least the first two (and perhaps the third) to be extensionally equivalent. Combining internalism about evidence with an externalist reliabilism about knowledge will drive the second and third properties apart, but both an across-the-board internalism and an across-the-board externalism plausibly allow for the possibility that the second and third properties are extensionally equivalent.

[43] I'm grateful to Kurt Sylvan for discussion of this.

epistemology that is a matter of how things are just with the agent's mind, and not with whether the external world cooperates. On the other hand, the dialectic can make the resistance of externalism at some stages but not others look unprincipled. This is so especially if one thinks of the motivation behind internalism as that of seeking a "luminous" epistemic property – that is, a property such that the believing agent is always in a position to know whether her states have that property. As Timothy Williamson, a very prominent epistemic externalist, has argued, it is doubtful whether there is *any* such property.[44] So if that is what internalism is trying to make good on, it is equally doubtful whether internalism about *any* epistemic notion can be vindicated.

However, not every internalist impulse need be driven by a concern for luminosity,[45] and I think that coherentism does an especially good job of illustrating a way of thinking about rationality as an internal matter without being motivated by any kind of concern for luminosity. According to the coherentist view of rationality, rationality is about lining up one's attitudes with one another in the right ways, and avoiding impermissible combinations of attitudes (or lack thereof). Whether one satisfies the relevant coherence requirements is clearly a matter of one's attitudes alone, then, and not of whether the external world co-operates. At the same time, there is no commitment whatsoever here to the idea that rationality is a luminous property. Although rationality is, on a coherentist view, a matter of whether one satisfies various coherence requirements, one may not always be in a position to know whether one satisfies these coherence requirements. For example, as Williamson suggests, one may be misled about one's own mental states.

---

[44] See Williamson (2000: chs. 4 and 8).

[45] Cf. Wedgwood (2002a).

Williamson's own "knowledge-first" epistemology understands notions like belief, evidence and justification in terms of knowledge rather than vice versa. On some readings, this has the extreme externalist result that there are no justified false beliefs. Williamson generally avoids the term 'rationality', instead preferring to talk of justification, and I think this is no accident. While one can perhaps get in a frame of mind on which 'justified' is to be understood in terms of knowledge, such that no false beliefs can be justified (since not known), the account stretches credibility as an account of rationality in any recognizable sense of the term.[46] In a sense I think Williamson may be better viewed as a kind of myth theorist about rationality,[47] denying that there is any serviceable or useful normative notion in the neighborhood of the ordinary concept of rationality (since that ordinary concept is not explicable in terms of knowledge).

But again, I think that the coherentist account shows that there is a perfectly serviceable notion of rationality that is clearly not to be understood in terms of knowledge, nor vulnerable to the kinds of arguments that might drive us towards externalism or "knowledge-first-ing" about other notions like evidence or justification. As long as one construes the relevant normative claims as narrow-scope ones – of the form: if X, then one should believe/is justified in believing/has decisive evidence for p – then Williamson's arguments that the relevant 'X' should be understood in terms of one's knowledge have the potential to get a grip. Mere belief certainly doesn't have that kind of fixed authority, after all. But as I explained in section 0.3, once one sees the wide-scoping option, we have requirements that are more symmetric than this. No individual attitude is being given authority over any

---

[46] Indeed, Littlejohn (2012), who defends the view that there are no justified false beliefs, is sure to distinguish justification from rationality and admit the possibility of *rational* false beliefs.

[47] Cf. the "myth theories" of Kolodny (2005, 2007, 2008a, 2008b) and Raz (2011: ch. 11). These views are, of course, importantly different from Williamson's view in a number of ways.

other attitude. The idea is simply that there are certain attitudes (some doxastic, others not) that are irrational *in combination with one another*. Whether such attitudes (when they are doxastic) amount to knowledge does not bear on the incoherence of such combinations.

In Williamson's resolutely externalist, knowledge-first epistemology, the distinctive incoherence of such combinations gets lost.[48] So even if one is moved by a Williamsonian approach to knowledge, evidence and justification, one can see the present dissertation as filling in an important part of the landscape of doxastic normativity that Williamson's approach leaves out. In fact, recognizing the distinctiveness of coherence requirements allows one, if one is a Williamsonian, to let one's externalism about evidence and knowledge run as wild as one likes, while still preserving an important dimension of what the internalist wanted – all without any problematic luminosity assumptions.

0.8   The plan for the dissertation

I will proceed as follows.

Chapter 1 takes up the project of showing that coherence requirements and evidential reasons can come into conflict. I use the possibility of such conflicts to argue that the two are distinct normative phenomena that cannot be reduced to one another, or lumped under the same general heading. In light of this, I argue for restricting the term 'rationality' to the satisfaction of coherence requirements only. Much as this is revisionary practice, the coherence requirement that I argue comes into conflict with evidential reasons is one that epistemologists have (rightly) wanted to affirm and account for, and one that is at the core of rational doxastic

---

[48] This is particularly evident in Williamson (2011), where he argues that one can know p even while the probability on one's evidence that one knows p is arbitrarily low. The conclusion in his knowledge-first framework seems to be that in such a case, it is absolutely fine to believe p while also taking it to be overwhelmingly probable that one doesn't know p. How these two mental states fit together just doesn't seem to be something Williamson considers.

agency. So if satisfaction of it is incompatible with believing what one's evidence supports, I argue, the claim that rationality requires one to believe what one's evidence supports is what should go. But, as I have said, even if one rejects my proposal about rationality (and 'rationality'), the substantive point about the metaphysical distinctness of coherence requirements and evidential reasons, and the possibility of conflict between them, remains.

The particular kind of conflict that I argue for arises from the possibility of all-things-considered misleading higher-order evidence. In the course of arguing for the possibility of conflicts between coherence requirements and evidential reasons, I will argue that allowing for such conflicts (and, as a precondition of that, making the distinction between coherence requirements and evidential reasons) helps us to resolve a puzzle in the literature about misleading higher-order evidence. So as well as illuminating the nature of coherence requirements, chapter 1 also provides an illustration of the epistemological fruitfulness of attending to them, and keeping them adequately distinct from claims about evidential reasons.

The conflicts illustrated in chapter 1, though they arise in the doxastic case, should also be of interest to theorists of practical rationality. Even though many theorists distinguish coherence requirements from reasons in the practical rationality literature, many take implicit or explicit views about the relationships between the two that preclude such conflicts.[49] That can make it look as if the debate between coherentist and reasons-responsiveness accounts is just one over whether to use 'rationality' in a more or a less demanding way. And it also suggests that the reasons-responsiveness theorist still gets to say that instances of incoherence are irrational. If there are such conflicts, these assumptions are false.

---

[49] Strikingly, this includes those with extremely divergent assumptions about rational requirements, including wide-scope coherentists (e.g. Dancy (1977), Darwall (1983), and at least the early Broome (1999, 2004)), reasons-responsiveness theorists (e.g. Lord 2014b), and "myth theorists" (e.g. Kolodny 2005, 2007, 2008a, 2008b).

Chapter 2 undertakes a more general defense of the coherentist view of rationality. I argue that "mere" coherence is an important, philosophically interesting notion, and say more about what it is and what makes coherence requirements distinctive. I also argue that coherence is central to our ordinary concept of rationality, in a way that actively tells against the reasons-responsiveness account. This last point is brought out mainly with respect to the case of actions and moral reasons. But my ultimate interest here is with beliefs and epistemic reasons. So I argue systematically that the two cases should be treated analogously. In the course of chapter 2, then, we get an argument for an important kind of parallel between the two spheres. I then offer an error theory to explain why our intuitions about rationality at least seem to depart from the coherentist picture in the epistemic case specifically. This part of the dissertation amounts to my most sustained interrogation of the chasm that I laid out in section 0.1, investigation of what might have caused it, and case against this kind of bifurcation in our theory of rationality. Finally, I ask whether there are *modified* versions of the reasons-responsiveness account to can do a better job of providing a unified account, and argue that such modified versions of the view fail, in a way that actually illustrates the appeal and power of the coherentist picture.

Chapter 3 returns to the issue about the scope of rational requirements that I introduced in section 0.3. As I made clear there, my view is that the fundamental coherence requirements are wide-scope. Many other philosophers who have preferred wide-scope requirements to narrow-scope requirements have also endorsed a corresponding *semantic* claim, namely that ordinary talk about rationality, despite appearances to the contrary, expresses wide-scope claims. Such a claim may appear necessary to avoid attributing massive error to ordinary speakers. However, it is becoming increasingly clear that the wide-scope *semantics* inadequately captures the meaning of ordinary talk about rationality. It seems, then,

that we are left with a dilemma: either give up the view that requirements of rationality are wide-scope, or accept an implausible semantics for ordinary talk about rationality, or attribute massive error to speakers. In chapter 3, I argue that this dilemma is only apparent, since we can appeal to a standard kind of contextualist semantics for modals to explain why narrow-scope *talk* comes out true *in virtue of* the wide-scope requirements.

My view, then, combines wide-scoping about the explanatorily fundamental requirements of rationality with a contextualist variant of a narrow-scope semantics. I argue that this view gives us the best of both worlds, as well as solving related puzzles and challenges for the extant views in the literature. As well as helping to precisify the account of the nature of coherence requirements that I am defending, the view developed in this chapter also bolsters the coherentist view of rationality advanced in this dissertation more generally, by showing how it can be made consistent with, and even explain, the intelligibility and truth of ordinary talk about rationality that it not on its face about *combinations* of attitudes – appearing, rather, to be about the rationality of *individual* attitudes.

In **chapter 4**, I bring the framework developed earlier in the dissertation to bear on the classic puzzle of why it is so hard to believe something at will – that is, directly and consciously on the basis of pragmatic considerations. I argue for three, highly interrelated, contentions. First, I argue that philosophers have tended to overstate the generality of the claim that believing at will is difficult. Second, I aim to cast doubt on the popular thought that the difficulty of believing at will can be explained by facts about substantive norms or reasons for belief. Third, I suggest that we can do better by focusing on coherence requirements. I argue that believing at will, in the paradigm cases, would require one to be incoherent in a way that is transparent to oneself – and, picking up on a more general feature of coherence requirements identified in chapter 2 – that this kind of transparent incoherence is hard for

47

agents to sustain. Believing at will thus serves as a case study illustrating the theoretical fruitfulness for epistemology of distinguishing substantive normative reasons from coherence requirements, and how the latter are poised to play *explanatory* roles that the former are not.

In **chapter 5**, I turn to putative deductive constraints such as deductive closure and consistency. One might think that, on a coherentist account, deductive logical constraints would be central to the requirements of rationality. However, there are very strong objections to deductive constraints from preface paradox-type risk aggregation cases. If preface paradox-type cases were simply understood as ones in which one's evidence supports a set of beliefs which fail to be deductively closed under logical implication, then one might think that this is another case of conflict between rational requirements on one hand and responding to reasons on the other. However, I show that one can motivate the preface paradox-type thought that having deductively closed beliefs is rationally impermissible in preface-type cases within a coherentist framework, without any appeal to evidence. Along the way, I defend some coherence requirements that govern the relationship between one's outright beliefs and one's credences, and examine how the rejection of deductive closure and consistency bears upon some questions in the descriptive theory of belief. However, I accept that there may be some weaker deductive constraints that still hold.

Finally, in the conclusion, I briefly look at some other places in epistemology where distinguishing coherence requirements and substantive reasons may help us to make progress, looking ahead to future work. In particular, I suggest that it may help in debates about rules for belief revision such as conditionalization and requirements governing responses to peer disagreement.

Throughout, I have two higher-level theoretical aims. First, I am engaging in the kind of enumeration of coherence requirements that I think needs to be done in order to make the

theory adequately determinate and compelling. I elaborate a number of individual coherence requirements as the dissertation proceeds. The coherence account embraces the idea that there may be many requirements of rationality on belief. So I do not claim that the limited number of wide-scope principles I endorse in this dissertation are the only wide-scope requirements of rationality on belief. In fact, I believe that there are probably many more. I remark on the method for uncovering them in section 2.1.

I have focused on the principles that I do because they seem to me important, interesting, and of central relevance to current epistemological debates, and because the space I have is ultimately limited. I hope they at least show by example how the coherentist approach, far from limiting rational requirements to a parochial role, is both theoretically fruitful and explanatorily powerful. But there are, I believe, still numerous further ways in which to apply it. In particular, most of the requirements that I explore here are requirements on outright attitudes (belief, disbelief, and suspension of judgment), as opposed to graded belief or credence. I will, in chapter 5, say something about some coherence requirements that govern the relationship between one's outright attitudes and one's credences. But I do not explore in depth the many putative requirements that govern relationships between credences (including probabilism, conditionalization, calibration requirements, and so on). In the conclusion, I say something about how the account developed here might be able to shed light on debates about these matters. But time and space have not permitted me to develop this here; I hope to be able to do so in future work.

My second higher-level theoretical aim is to address the aforementioned concern that coherence requirements are not distinctively epistemologically interesting. A reader may worry that if rationality is to be understood as coherence, then epistemology turns out not really to be about rationality at all. So I aim to show that coherence requirements really do have

application to important epistemological debates. Indeed, I suggest that these debates have, in general, been distorted and misunderstood through failure to see the option of understanding certain central epistemic norms as wide-scope coherence requirements of rationality. Ultimately, I attempt to "show by doing" the explanatory power of the coherence approach; in this way, the theory earns its keep. Understood slightly differently, the aim is to show how much in epistemology we can achieve and explain with an account of rational belief that utilizes coherence requirements alone.

Chapters

Chapter 1
Rational Belief without Evidence-Responsiveness

1.1   A puzzle

Consider the following three claims. Throughout, D(p) is a possible doxastic attitude (D) of a subject (S) towards a proposition (p): believing p, disbelieving p, or suspending judgment about p.[50]

Evidence requirement (ER). If S's evidence supports D(p), then rationality requires of S that she takes D(p),

Inter-level coherence (ILC). Rationality requires of S that

(i)   S believes that her evidence supports D(p) → she takes D(p)

(ii)   S believes that her evidence does not support D(p) → she does not take D(p)

Possibility of iterative failure (PIF). It is possible that:

(i)   S's evidence supports D(p); and

(ii)   S's evidence supports believing that her evidence does not support

D(p)

---

[50] So our focus here is on outright attitudes, and not on graded attitudes (i.e., credences). There are analogues of both (ER) and (ILC) for credences. But formulating the credence-specific version of (ILC) turns out to be very complex and to raise a host of its own issues. So to keep things manageable, I focus on outright attitudes here.

(ER) is the claim that rationality requires one to take the attitudes that one's evidence supports. (ILC) is a requirement of coherence between one's first-order beliefs and one's higher-order beliefs. (PIF) is the claim that one can have all-things-considered misleading higher-order evidence, such that one's evidence supports a false claim about what one's evidence supports (I call such cases of "iterative failures", since they are cases where one's evidence fails to iterate across levels of belief).

Suppose now that (PIF) obtains with respect to some doxastic attitude D towards some proposition q: that is, S's evidence supports D(q), but S's evidence supports believing that her evidence does not support D(q). Then, S cannot satisfy both (ER) and (ILC). To see that, consider this table:

	S takes D(q)	**S doesn't take D(q)**
S believes that her evidence does not support D(q)	S violates (ILC)	S violates (ER)
S does not believe that her evidence does not support D(q)	S violates (ER)	S violates (ER) (twice over)

Table 1: our puzzle

Either S takes D(q), or she doesn't. And either S believes that her evidence does not support D(q), or she doesn't. So S must be in one of the four boxes. Yet in each of the four boxes, S violates either (ER) or (ILC). So if there are cases of iterative failure – as (PIF) says there are – then by the same token, there are cases where (ER) and (ILC) are not co-satisfiable.

Something like this puzzle – not quite in the form that I have spelled it out in here – has been noticed before. As far as I can see, the existing responses in the literature fall into two broad categories:

1. Deny that cases of iterative failure are possible (that is, deny (PIF)). So conflicts between (ER) and (ILC) cannot arise, at least not in this way[51]

2. Deny that (ILC) is a genuine requirement of rationality, and hold that in cases of iterative failure, one can violate (ILC) while remaining rationally flawless[52]

I am going to argue that both of these responses fail. (ILC) and (PIF) are both independently plausible, I will argue, and neither provides principled reason for denying the other. Most philosophers who discuss our puzzle seem to assume that options 1 and 2 are the only options. If this assumption were true, such that we simply must deny either (ILC) or (PIF), then the truth of one would of course then provide very strong grounds indeed for denying the other. However, the assumption is too hasty. There are at least two further options available to us. The first (option 3 overall), not unequivocally endorsed by anyone in response to this particular puzzle, but in the spirit of some remarks made by at least one philosopher,[53] is:

3. Maintain that (ER) and (ILC) are both requirements of rationality (in the same sense of 'rationality'), and hold that cases of iterative failure are rational dilemmas, whereby, whatever one does, one violates a requirement of rationality

---

[51] Feldman (2005); White (2007: 120); Huemer (2011); Greco (2014a); Titelbaum (2015b). Horowitz (2014) takes a nuanced stance on (ILC) – arguing that it holds for the most part, but occasionally fails. Nevertheless, like the other authors mentioned here, she assumes that any admission of (PIF) requires admitting a corresponding failure of (ILC).

[52] Coates (2012); Lasonen-Aarnio (2015); Weatherson (ms.); and arguably Wedgwood (2012a). In some ways this strategy is inspired by the approach taken by Alston (1980), who introduced much of the talk about different "levels" in epistemology and skepticism about whether facts about evidence or justification will always line up across levels. However, nothing Alston says explicitly commits him to option 2, or even to (PIF) itself.

[53] Christensen (2007, 2010, 2013).

The last option, and the one I favor, is:

4. Maintain that (ER) and (ILC) are, properly understood, different kinds of normative claim, such that they should not be stated using the same normative concept.[54]

One way to execute this strategy (and the way that I prefer) is to say that only (ILC), and other coherence requirements like it, are requirements of *rationality*, strictly speaking. As such, we should reject (ER) as it is currently stated, and replace it with:

(ER*) If S's evidence supports D(p), then S has most epistemic reason to take D(p)

Of course, there are different ways to use the relevant terminology here. The most important feature of strategy (4) is that it affirms that (ER) and (ILC) are different kinds of normative claim.

I want to stress at the outset that there are a number of commonalities between option 3 and option 4. Perhaps most importantly, both strategies accept that (genuine) coherence requirements can come into conflict with evidence-responsiveness. The difference between the two strategies is that while option 3 construes this as a conflict between two different requirements of the fundamentally same kind (two different requirements of rationality), option 4 construes it as a conflict between two fundamentally different kinds of normative demand (on my preferred terminology, that of satisfying the requirements of rationality, and responding to one's reasons). Later, I will offer some reasons to prefer option 4 over strategy

---

[54] This option does not, as far as I can see, even feature in existing taxonomies of the options in response to our puzzles and related ones. Compare, e.g., Huemer (2011: 5); Greco (2014a: 206); Titelbaum (2015b: 278-9).

3. However, on *either* view, it's crucial to see that the following popular claim (alluded to in section 0.1) turns out to be false:

Evidentialism. S's doxastic states are rational iff they satisfy (ER).

On option 3, Evidentialism is false because satisfying (ER) is not sufficient for being rational. This is because satisfying (ER) does not suffice for satisfying (ILC), and (ILC) is a genuine requirement of rationality. On option 4, Evidentialism is false because satisfying (ER) is neither necessary nor sufficient for being rational (properly speaking). So, on both views, rationality is not *simply* a matter of taking the doxastic attitudes that one's evidence supports. The idea that evidence-responsiveness guarantees coherence, as well as the bolder idea that coherence requirements are ultimately reducible to claims about evidence, both fail.

The plan of attack for the remainder of this chapter is as follows. In section 1.2, I will issue some preliminary clarifications about how the notion of evidential support explained in section 0.6 above plugs in to my argument. In section 1.3, I will give an explication and partial defense of (ILC). In section 1.4, I will give an explication and defense of (PIF). Having argued that cases of iterative failure are possible, I will in section 1.5 complete my defense of (ILC) by arguing that its plausibility is not diminished in cases of iterative failure specifically. That will complete my case that (genuine) coherence requirements can come into conflict with evidence-responsiveness. Sections 1.6 and 1.7 turn to the issues that remain once we have already accepted that. Section 1.6 argues for option 4 over option 3 – that is, in favor of distinguishing the kinds of normative demands in play, rather than describing the situation as one where demands on the *same* kind come into conflict. Section 1.7 argues for my particular implementation of option 4: that of reserving the term 'rationality' for satisfaction of

coherence requirements, and framing the demands of evidence-responsiveness in terms of *reasons*.

1.2   A bit more on evidential support

As I said in section 0.6, I am using the term 'support' in the *pro toto* sense, such that what the evidence supports is the doxastic attitude that one has *most* reason to take, given one's total evidence. If the evidence *pro tanto* supports the different possible attitudes to various degrees, there must be one (or more) attitudes that it supports the most. So, given my usage, there will always be at least one attitude that the evidence supports with respect to a proposition. If your evidence does not support believing p or support disbelieving p, then it supports suspending judgment about p.[55] So, for example, a case where you have very scant evidence both for and against some proposition is not a case where your total evidence doesn't support *any* doxastic attitude; rather, it is one where your total evidence supports suspending judgment. Making this clear is one advantage of talking of evidential support for attitudes, rather than propositions.

What about the possibility that your evidence supports more than one different doxastic attitude equally well?[56] My usage of 'support' does not exclude this possibility, since the reasons for taking two or more different attitudes might be equal. In such a case, there is a tie in terms of what attitude the evidence gives one most reason to take. Anyone who is

---

[55] Given this usage, (ER) entails the claim that if S's evidence does not support D(p), then rationality requires of S that she does not take D(p). The entailment runs in both directions for cases in which a subject must either believe, disbelieve, or suspend judgment, so if there are some such cases, the argument will also extend to this more negative analogue of (ER).

[56] This may put readers in mind of the debate between "uniqueness" and "permissivism": see, e.g., White (2005); Ballantyne & Coffman (2011); Schoenfield (2014). However, these authors tend to treat the thesis that there is always one unique attitude that the evidence supports interchangeably with the thesis that there is always one unique attitude that is *rational* given a subject's evidence. These two formulations are only equivalent if we assume that rationality is a matter of evidence-responsiveness. I reject uniqueness as a thesis about rationality, but am agnostic on it as a thesis about what the evidence supports.

attracted to this possibility, however, will not like (ER) as it is currently formulated, since we definitely don't want to say that rationality requires you to take *each* of the doxastic attitudes that your evidence supports equally well. Rather, those who think this situation is possible should read D (as it occurs in every principle that I refer to) as the disjunction of the different attitudes that the evidence supports. My arguments will be unaffected by this modification.[57]

1.3   (ILC) explicated and defended

Here, once again, is the inter-level coherence requirement:

Inter-level coherence (ILC). Rationality requires of S that

(iii)   S believes that her evidence supports $D(p) \rightarrow$ she takes $D(p)$

(iv)   S believes that her evidence does not support $D(p) \rightarrow$ she does not take $D(p)$

As the name suggests, (ILC) is a (putative) coherence requirement. It is about the relationship between your first-order attitudes and your higher-order attitudes, and specifies particular combinations of first-order attitudes (or lack of) and higher-order attitudes that are irrational. Stipulatively, '$\rightarrow$' represents the material conditional.[58] If you don't like material conditionals, you can convert (iii) and (iv) into the equivalent disjunctions. So, for example, (iii) becomes:

---

[57] (ER) may in a further way seem too strong even for its sympathizers, since it appears to require that you have beliefs in propositions which your evidence supports but you don't care about at all. If one is moved by this problem, one could restrict (ER) to propositions which one has some doxastic attitude towards, or even explicitly to propositions that one cares about. These modifications would also leave my main arguments unaffected.

[58] This is to make no commitment that the material conditional is expressed in ordinary conditional *talk* about rationality. See chapter 3 below.

"S takes D(p) ∨ ¬S believes that her evidence supports D(p)." What is forbidden is S simultaneously believing that her evidence supports D(p) but not taking D(p).

Relatedly, (ILC) is wide-scope. This is important. It may be tempting to think of (ILC) as a kind of "subjectivization" of (ER), replacing the notion that rationality is about conforming your beliefs to the evidence with the idea that rationality is about conforming your beliefs to what you believe to be the evidence.[59] But in an important way this is misleading.[60] (ILC) is not the following:

Narrow scope inter-level coherence (NILC).

(i)   S believes that her evidence supports D(p) → rationality requires of S that she takes D(p)

(ii)   S believes that her evidence does not support D(p) → rationality requires of S that she does not take D(p)

As least read naively, this narrow scope requirement is implausible.[61] It entails that, if you have crazy second-order beliefs about what your evidence supports, then rationality *requires* you to have correspondingly crazy first-order beliefs. Moreover, (NILC) is obviously incompatible with (ER). For in any situation where someone has mistaken beliefs about what their evidence supports, (ER) and (NILC) will issue contradictory requirements. A fan of (ER) will feel principled in rejecting (NILC).

---

[59] Cf. the criticism of (ILC) in Wedgwood (2012a: 291); see also White (2007).

[60] See also Feldman (1988: 412); Huemer (2011: 3).

[61] I allow that (NILC) may be true given a particular, contextually restricted reading of 'rationality requires', albeit only in virtue of the fundamental truth of the wide-scope requirement (ILC). This restricted reading does not have the bad consequences that the naïve reading has. See section 3.6 below.

By contrast, (ILC) does not have these same consequences. Because (ILC) is wide-scope, it is neutral on *how* to resolve inter-level mismatches. Consequently, it does not give any special authority to one's beliefs about one's evidence. They are just as open to revision as one's first-order beliefs. Suppose you violate (ILC), so that there is a mismatch between your first-order attitudes and your higher-order beliefs about what your evidence supports. You can come to satisfy (ILC) *either* by revising your first-order attitude, or by revising your higher-order belief. (ILC) does not require either revision in particular, and leaves it open that in some particular case you ought (for other reasons) to satisfy it by one kind of revision rather than the other kind. All it says is that you are required to revise *something* so as to eliminate the conflict in your doxastic states. Consequently, one might think initially that (ER) and (ILC) will always be co-satisfiable. The fan of (ER) can hold that when one has mistaken beliefs about what one's evidence supports, one should come to satisfy both (ER) and (ILC) by revising one's beliefs about what the evidence supports.[62] As we will see, it is only when we bring in (PIF) that the two putative requirements come into conflict.

Note also that (ILC) stays silent about situations in which you suspend judgment about what your evidence supports. So, (ILC) does not say that it is irrational to believe p while suspending judgment about whether one's evidence supports believing p.[63] This is, I think, as it should be. It is important to distinguish two very different states here: suspending judgment about whether one's evidence supports believing, and believing that one's evidence supports suspending judgment. In the latter case, (ILC) does forbid one from believing. When one believes that one's evidence supports suspending judgment, one believes that it supports

[62] Again, see Feldman (1988: 412).

[63] Here I agree with Hazlett (2012), and disagree with Horowitz (2014), Feldman (2005: 118 n. 6); Huemer (2011: 1), and Adler (2002b: 311 n. 6).

suspending judgment *rather* than believing; that it does not support believing. And so it would violate (ILC) to believe in such a case.

In the former case, however, (ILC) does not forbid one from believing. In this case, one is unsure about whether one's evidence supports believing, or whether it supports some other attitude such as suspending judgment.[64] Given that one is just as unsure about whether one's evidence supports suspending judgment as one is about whether one's evidence supports believing, it is no more incoherent with one's higher-order state to believe than it is to suspend judgment. So I do not think that the correct formulation of (ILC) should forbid one from believing in such a case. Nor does it require (even more strongly) that whenever you have a doxastic attitude, you have a corresponding higher-order belief that the evidence supports this doxastic attitude; you may have no higher-order attitude at all.[65]

These clarifications help us to see what the motivation for (ILC) is not. Cases of iterative failure make it clear that one can be blamelessly misled about what one's evidence supports. One might think that (ILC) is motivated by an (extreme) internalism (recall section 0.7) that requires the normatively relevant property for the assessment of beliefs to be "luminous", in the sense that one *cannot* be blamelessly misled about it. The thought might be that although one can be blamelessly misled about what one's evidence supports, one cannot be blamelessly misled about what one believes one's evidence supports. So one should replace (ER), which refers to what one's evidence supports, with (ILC), which refers to what one

---

[64] It's important here that what one is unsure about is which *attitude* the evidence supports, such that one is unsure about whether the evidence supports *suspending judgment* as well as being unsure about whether it supports believing. Again, if we talk of support for propositions instead of support for attitudes, we obscure matters here. We will then be liable to throw *both* what I am calling "the former case" and "the latter case" under the general heading of "cases where one is unsure about both whether the evidence supports p and whether the evidence supports not-p".

[65] Cf. the view criticized by Bergmann (2005).

believes that one's evidence supports. But this is not my thinking: in fact, I reject these claims. It is part of the spirit behind (PIF) that one can be misled about just about anything interesting, including one's own beliefs.[66] In my view, to say that one can be misled about one's evidence, but not about one's beliefs about one's evidence, represents a kind of unstable midway position.

(ILC) – or variants of it close enough to preserve the validity of my argument – is widely accepted.[67] So if (ILC) does not rely on any extreme internalism, what does motivate it? Let's begin with Adam Elga's (2005: 115) presentation of a case where (ILC) is violated, and his reaction:

> "My friend Daria believed in astrology. For example, she thought that because of her astrological sign she was going to be particularly lucky over the next few weeks. That was bad enough. But when I tried to persuade her that astrology is unfounded, I discovered something even worse. I gave Daria evidence against astrology — studies showing that the position of the distant stars at the time of one's birth has no bearing on one's personality or prospects. Daria agreed that the studies were significant evidence against the truth of astrology, and that she had no countervailing evidence of comparable strength. But that was not the end of the matter. "I still believe in astrology just as much as I did before seeing the studies," she said. "Believing in astrology makes me happy."
>
> I was floored. Daria's original belief in astrology was less than perfectly reasonable. But this—believing in astrology even though by her own lights the evidence went against it—was an insult to rationality."

---

[66] See, amongst others, Williamson (2000: esp. ch. 4); Schwitzgebel (2008); Srinivasan (2015). For an application of these points to draw lessons about norms epistemic and otherwise, see Hawthorne & Srinivasan (2013); Srinivasan (forthcoming).

[67] G. Harman (1986: 39); Bergmann (2005); Elga (2005); Feldman (2005); Kolodny (2005: 521); Scanlon (2007); Christensen (2007, 2013); Smithies (2012: 283-4) Broome (2013: 98); Greco (2014a); Horowitz (2014); Titelbaum (2015b).

Notice that Elga here takes (ILC) to be a more fundamental and non-negotiable part of rationality than (ER). In calling Daria's original belief "less than perfectly reasonable," he stops short of calling it *irrational* – it is her violation of (ILC) that is, for Elga, "an insult to rationality."

Why think this? One way to motivate the thought is to appeal to an analogy with the so-called "enkratic requirement" of practical rationality. This requirement forbids instances of akrasia, where you fail to intend to do what you believe you ought to do.[68] Several writers who defend something close to (ILC) have presented it as a requirement that forbids "epistemic akrasia."[69]

At least as I have formulated it, (ILC) is not quite an anti-akrasia requirement, strictly speaking, because it refers to what one believes one's evidence supports, and not to what one believes one ought to believe.[70] However, as I said earlier, what one's evidence supports is to be understood in terms of what one's evidence gives one most reason to believe. And reasons and 'ought' are plausibly connected in some systematic way (as I said in section 0.6). So the analogy is still quite close. The idea here if what one has most (epistemic) *reason* to believe is constituted by what one's evidence supports, then violating (ILC) effectively commits one to regarding one's own belief as unreasonable. And to be in such a state is to be failing by one's own lights in a way that makes one incoherent. This is why (ILC) is a genuine requirement of coherence.

---

[68] See Broome (2013).

[69] See Horowitz (2014); Greco (2014a); Titelbaum (2015b).

[70] Formulations vary here: Greco's requirement refers to what one believes what ought to believe; Titelbaum's to what one believes one is rational to believe; Horowitz's – like mine – to what one's evidence supports. Greco suggests that these formulations are interchangeable, but I am not convinced – I'll come back to this point at the end. Broome (2013: 94-96) puts a powerful objection to a formulation in terms of 'ought' or reasons, but accepts a version referring to what one's evidence "shows" (*ibid*.: 98).

Note that this way of motivating (ILC) actually, far from denying the normative importance of what one's evidence actually supports (as a simple "subjectivization" of (ER) might be thought to), actually *trades* on this normative importance. As such, (ILC) should be attractive to those who appreciate the normative importance of evidence: violating (ILC) amounts to what is, from the subject's own point of view, an act of disrespect for her evidence.[71] That is one reason why (ILC) has actually been accepted by epistemologists that take evidence to be normatively fundamental.[72]

At the same time, these observations already start to point us some of the salient differences between reasons and coherence requirements. What one has most (epistemic) reason to believe is just constituted by what one's actual evidence supports. But what involves the agent in *incoherence* is that she *takes herself* to be failing to believe what her evidence supports. These facts are, at least arguably, related; more on this in chapter 4. Nevertheless, if we confuse coherence requirements and reasons, we muddle these two separate phenomena. The idea is not that what one *takes* to be one's evidence is determinative of what one has *reason* to believe, any more than the enkratic requirement claims that what you believe you ought to do is determinative of what you ought to do. Nor is the idea that there is really some kind of internal incoherence in failing to believe what one's evidence *actually* supports. Rather, reasons and coherence requirements have to be pulled apart here. Much more on that in sections 1.6 and 1.7.

Some are unmoved by the appeal to analogy with the enkratic requirement. They may reject that requirement. Fortunately, there is another, independent way to motivate (ILC) that

---

[71] This is, I take it, why Feldman (2005) says that agents who satisfy (ILC) "respect their evidence". However, it would be better if Feldman made explicit the difference between this kind of respect for what one *takes* the evidence to support from responsiveness to what one's evidence *actually* supports.

[72] Many if not all of those cited as supporting (ILC) in fn. 67 are in this category,

does not appeal to the analogy with the enkratic requirement (and may put the former on a stronger footing than the latter). Consider what it is for some body of evidence to support some belief. It seems overwhelmingly plausible that it is constitutive of such evidential support that when some body of evidence E lends support to believing p, E in some sense makes p more likely to be true. It's part of being competent with the concept of evidence that you when you judge some piece of evidence to lend *pro tanto* support to believing p, you take it that this increases the likelihood that p is true.

But clearly, there are certain possible incoherencies – of the kind that amount to irrationality – between judgments about whether p is true on one hand, and judgments about whether p is likely to be true on the other. It is clearly incoherent to maintain, for example, that p is true, while also judging that p is not likely to be true. It may not be strictly inconsistent, but it is Moore-paradoxical.[73] Moore-paradoxical combinations of attitudes are paradigm instances of attitudes that do not fit together rationally.

Of course, there are different kinds of likelihood, some of which don't generate the Moorean incoherence. I can believe that I share a birthday with my partner whilst also believing that my sharing a birthday with her is not likely in the sense that the prior probability of two randomly sampled individuals from the general population sharing the same birthday is fairly low. That probability, however, is a kind of background statistical frequency given a restricted set of information that is more impoverished than my actual evidence base. For the purposes of this probability, I exclude pertinent evidence that I have about what our birthdays actually are. The notion of likelihood on which it is incoherent to believe that p is true and to

---

[73] See also Adler (2002), Feldman (2005), Huemer (2011) and Smithies (2012).

believe that p is not likely to be true, then, is that of likelihood *on one's current evidence*.[74] But for a proposition to evidentially support a belief just is for it to be make the proposition believing more evidentially likely. So, just as it is incoherent to judge that p is true while judging it not to be likely (on one's evidence) to be true, it is likewise incoherent to judge that p is true while judging it not to be sufficiently supported by one's evidence. And that gets us to the idea that violating (ILC) – or, at least, part (ii) of (ILC) – is incoherent.[75]

What are the bounds of the vague term 'likely' here? It's not that there is some particular level of likelihood, L, such that that rationality requires that if one judges the likelihood of p to be greater than or equal to L, one believes p. Rather, more liberally, rationality requires that if one oneself judges the likelihood of p to be sufficiently high such that believing p is conclusively supported – that is, to be the attitude that one's evidence gives one most reason to take – then one believes p. And rationality requires that if one judges the likelihood of p to be insufficiently high for believing p, then one does not believe p.

Very recently, some authors have tried to put pressure on principles like (ILC). The primary way in which this has been done, however, is to produce examples of iterative failure, where one has misleading evidence about what one's evidence supports, and assume that they show that principles like (ILC) fail.[76] As I have already suggested, it is too hasty to just assume

---

[74] This is the more natural sense of likelihood to be expressed by the phrase 'it is likely that p'. If someone asks me, "how likely is it that you'll have a joint birthday party this year?," I will not answer "very unlikely, since it is unlikely that we share the same birthday, even though we as a matter of fact do." Usually, unless context makes it clear otherwise, 'it is (un)likely that p' makes use of all the pertinent information about whether p. To exclude information, we explicitly draw attention to some narrower body of information and/or use the past tense. So I might say, "given what we knew when we met each other, it wasn't very likely that we would turn out to have the same birthday". It's more natural to say this that to just flatly say, "it's unlikely that we share a birthday", if you know that you actually do share a birthday.

[75] An analogous story can be given for part (i). I have focused on part (ii) because it is the part of (ILC) that is used in my argument here (cf. section 1.1).

[76] See Coates (2012); Lasonen-Aarnio (2015); Weatherson (ms.).

that any instance of (PIF) is a failure of (ILC): that rests upon the assumption that coherence requirements cannot come into conflict with the demands of evidence-responsiveness, or worse, upon conflating the two. However, what *would* be helpful for the denier of (ILC) would be if she could show that in cases of iterative failure, the usual positive rationale for accepting (ILC) – the one just sketched – fails to apply. We will only be able to assess that once we have concrete cases of iterative failure on the table. So I will return to it in section 1.5.

A more independent way of trying to put pressure on (ILC) might be to appeal to special cases where individuals purportedly take themselves to believe, and to be warranted in believing, without evidence.[77] One example might be religious belief; another might be anti-skeptical "hinge propositions" that one takes for granted such as "in general, things are roughly as they appear to the senses" or "I am not a brain-in-a-vat."[78] The thought would be that one can rationally believe these things whilst acknowledging that one has no good evidence for them. The two cases have something in common; both seem to be something like a foundational "leap of faith" from the agent's point of view. That's not to say that everyone who believes in God or who believes anti-skeptical hinge propositions is taking (or need think of themselves as taking) a leap of faith. Some people may simply take themselves to have good evidence for these propositions in a relatively straightforward manner not that different from their other beliefs. But at least some individuals seem to experience these beliefs as leaps of faith in some sense.

These examples raise huge issues that I cannot begin to deal with adequately here, but I will make a few remarks about their bearing on (ILC). Most importantly, the diagnosis of the

---

[77] Thanks to Keith DeRose for pushing this point with me.

[78] See, amongst others, Plantinga (1981) on the former case and DeRose (2000) on the latter.

person who takes herself to be taking a leap of faith as violating (ILC) is not at all obvious. There are, in fact, several ways to interpret, understand or develop the case, not all of which involve a violation of (ILC):

(i)   It may be that she believes the first-order proposition p while simply *not believing* herself to have conclusive evidence for believing p. For example, she may be unsure what to say about what the evidence supports when it comes to hinge propositions or to religious beliefs. But that isn't enough for her to violate (ILC): (ILC) only speaks to the case where one *does* have some higher-order judgment about one's evidence.

(ii)   It may be that she believes p while taking it that the evidence does not decide between the different possible doxastic attitudes toward p. She may think, for example that the evidence supports believing and suspending judgment (and perhaps also disbelieving) equally well. In light of her evidence, she might think, she has no better reason to believe than to suspend judgment; yet she might still believe. But this too isn't a violation of (ILC). As I said in section 1.2, the notion of an attitude being (*pro toto*) supported by the evidence in our sense is to be understood in terms of its being the attitude that the evidence gives one most reason to take, and this allows for ties. When you believe that the evidence supports some attitude tied-best, you do not count as believing that the evidence doesn't support this attitude in the sense we are interested in. This is so even if multiple attitudes are tied for best just because the evidence does not support any attitude particularly strongly, or even at all.

(iii)   It may actually be that she believes p yet takes it that believing p is not the attitude that her evidence gives her most reason to take. Then she does violate (ILC). Note that given the terminological clarification just given in the description of (ii), this commits her to thinking that there is some other attitude – perhaps suspending judgment, in the cases at hand – that the evidence supports better (though she might not be sure which attitude this is). So it will only be plausible to attribute this state to her if it is plausible to attribute this commitment.

(iv)   It may be that she does take it that believing is not the attitude that her evidence gives her most reason to take, but does not actually count as believing p. Perhaps her attitude is that of assumption, or faith, or acceptance, where these attitudes fall short of belief proper.[79] Then she does not violate (ILC).

(v)   Finally, it may actually be that she believes while actively judging that belief *is* the attitude that her evidence gives her most reason to take. Obviously, here she doesn't violate (ILC). How could this be experienced as a leap of faith, though? It might still seem to her that she is taking a simultaneous leap of faith *both* in believing and in believing that her evidence gives her most reason to believe. This can be illustrated best in the anti-skeptical case. Suppose that one is a (tacit?) externalist about evidence, and one thinks that whether one has good evidence for whether one is not a brain-in-a-vat (BIV) partially depends on whether one *actually is* a BIV or not.[80] If one is not a BIV, then things are set up reliably such that in light of one's ordinary sensory experiences, one has most evidential reason to

---

[79] Cf. Wright (2004); Sherman & Harman (2011); Buchak (2012: esp. fn. 2).

[80] Cf. Williamson (2000: ch. 8); Pritchard (2002).

believe that one has hands and that one is not a handless brain-in-a-vat. If one is a BIV, however, this sensory experience is unreliable and thus is not good evidence that one has hands; so one has most evidential reason to suspend judgment about whether one is a BIV. Now if one has this view of evidence, then believing the anti-skeptical hinge proposition that one isn't a brain in a vat will then naturally bring along with it the belief that one's evidence supports believing that one is not a BIV. So one assumes not just that one is not a BIV but that one has good evidence for this. Still, this double-whammy of assumptions can still feel like a leap of faith of sorts. Specifically, it may not be based on any *antecedent* judgment that one has evidence that one is not a BIV – "antecedent" not in a temporal sense but in the sense of its (apparent) status as evidence being independent of the claim that one actually isn't a BIV.

Only given diagnosis (iii) does the agent violate (ILC). But once (iii) is distinguished from (i), (ii), (iv) and (v), I myself do not feel any reluctance to say that the agent in case (iii) *is* irrational. As these other possibilities make clear, that does not mean that everyone who believes something on a "leap of faith" in some broad sense of the term is irrational. Moreover, no doubt the agent herself will often not have distinguished possibilities (i)-(v) clearly for herself, and so may describe herself in a way that sounds like one possibility when the best way of describing her mental state is really closer to a different one. Loose talk of "believing without evidence" does not adequately distinguish the above possibilities. And in some cases it might be somewhat indeterminate which of the above descriptions actually best captures the agent's mental states. Her rationality might then be correspondingly indeterminate.

Another point that must be borne in mind in interpreting these cases is that people sometimes speak with a rather restricted or narrow notion of 'evidence'. Sometimes people speak as if testimony, or private personal experience, or *a priori* considerations, cannot count as evidence. So then it might be that some religious believers (for example) describe themselves as "believing without evidence", even though they still (either explicitly or tacitly) take there to be weighty considerations that are strong indications of the truth of their religious belief, and in light of which they judge that belief to be extremely likely to be true. In such cases, I think the agent does (at least tacitly) take herself to have evidence at least in the sense of the term 'evidence' that is at issue in (ILC), as explicated above, and perhaps in any sense of the term that can be adequately made sense of. So I do not think that such an agent violates (ILC) either. Even if she describes herself in a way that makes it sound like she is in a case of type (iii), this may be an artifact of a narrow use of the term 'evidence'.[81]

Even with all this said, some may insist that the agent in case (iii) could be perfectly rational. I disagree. But it's worth noting that if one thinks this, and that is one's reason for denying (ILC), then one will probably also deny (ER). For the idea that leaps of faith of kind (iii) can be rational is presumably not that it could be OK to believe while taking oneself to lack adequate evidence but yet that it could not be OK to believe while *actually* lacking adequate evidence. If my suspicion is right here, then the person who objects to (ILC) on these grounds here would actually join me in rejecting (ER). Still, I have engaged this objection to (ILC)

---

[81] What if you think this narrower notion of evidence is the correct one? In this case, it sounds like you think there are epistemic reasons for belief that are in some important sense non-evidential. So you should understand (ILC) in terms of the notion of most epistemic reason rather than that of evidential support. That should still allow the argument of the chapter as a whole to go through. For you will likely allow that there are at least some cases where the evidence is decisive in determining what you have most epistemic reason to believe. So, if there can be iterative failures of evidence in those cases, the conflict between coherence and reasons-responsiveness that I am ultimately arguing for will still get onto the table.

because it is, in my view, a central requirement of rationality, and I want to use it for other purposes unconnected with the rejection of (ER) in chapter 4.

That brings my defense of (ILC) to a close for now. As I said, in section 1.5 we'll examine the thought that (ILC) breaks down specifically in cases of iterative failure. Before that, we should have the possibility of iterative failure (PIF) on the table.

1.4   (PIF) defended

You now may be wondering whether denying (PIF) might be the way to go. Here it is:

Possibility of iterative failure (PIF). It is possible that:

(iii)   S's evidence supports D(p); and

(iv)   S's evidence supports believing that her evidence does not support

D(p)

Denying (PIF) requires denying that one can have all-things considered misleading evidence about what one's evidence supports. In other words, justified false beliefs about what one's evidence supports are impossible. That is a strong claim, as any claim that a particular kind of justified false belief is impossible would be.

It's important to distinguish two quite different ways to be misled about what one's evidence supports. On one hand, there is the possibility that one might be misled about which items are part of one's body of evidence: that one might be led to think that some proposition (or fact, or appearance) is part of one's evidence when it actually isn't, or led to think that some proposition (or fact, or appearance) isn't part of one's evidence when it actually is. In this case, one is misled about what one's evidence *is*. On the other hand, there is the possibility

71

that one is misled not about what one's evidence is, but about the *evidential support relations* that hold between particular items in one's body of evidence and potential doxastic attitudes.

One can be misled about these support relations even if one identifies which items are *part* of one's evidence entirely correctly; similarly, one can be mistaken about which items are part of one's evidence even if one is omniscient about which (potential) items support which attitudes, and how strongly. Both of these ways of being misled will eventuate in being misled about what one's total body of evidence supports, and so to the possibility of iterative failure. But the source of the mistake will be different in the two cases: in one case it is traced to misleading evidence about what one's evidence *is* in the first place; in the other, to misleading evidence about the evidential support relations between one's evidence (even if that body of evidence is identified correct correctly) and potential doxastic attitudes.

In a moment, I'll take these cases one at a time, and try to explain how iterative failures of either kind can come about. Before that, a point about method. I will try throughout this section to give intuitive cases in which it is plausible that iterative failure occurs. However, there is a very general reason why it is difficult to give knockdown intuitive cases of such iterative failure, namely that it is always a somewhat vague and hard-to-identify matter just how much evidence is required to decisively support a doxastic attitude. Cases of iterative failure exploit the idea that sometimes your evidential support for a first-order attitude is stronger than the support for some corresponding higher-order attitude, or vice versa. Thus, in cases at the margins, it may be enough to *decisively* support the first-order attitude, without being enough to tip the balance of evidence at the level of higher-order attitudes. But because of the difficulty of determining how much evidence is decisive, it's always possible in any concrete case to protest that it strikes one that the evidence is not enough to decisively support the first-order attitude after all, or (perhaps) that it is enough to support the corresponding

higher-order attitude. (This same problem arises equally for the more familiar project of giving cases of the failure of the "KK" principle, which I also discuss below.) For these reasons, at times I will need to supplement my cases with theoretical arguments to the effect that we should not always expect things to evidentially line up across levels in the way that the opponent of iterative failure wants. The concrete cases thus serve to illustrate in what *kinds* of cases iterative failure may occur, even if one can quibble over the particular cases given.

(a) Cases of being misled about what one's evidence is

I'll begin by illustrating a method for generating cases of iterative failure out of failures of the so-called "KK principle", on which if one knows, then one knows that one knows. It shouldn't be surprising that these two possibilities are linked. If KK fails, the knowledge fails to iterate across levels; if that can happen with knowledge, it's not surprising that it can also happen with evidential support.

Let's begin with a classic case of KK failure, that of the unconfident examinee.[82] In this case, a high-school student, Sayeqa, is sitting a history exam, and is asked for the dates of various English monarchs' ascendance to the throne. For each monarch, Sayeqa gives the correct answer. Her doing so is based on her having learnt the dates in a normal way, through instruction in a classroom, and coming to associate the correct dates with each monarch. What she is in fact doing is recalling these dates based on this association. Yet she is of a nervous

---

[82] This case is first due to Radford (1966); what appears here is my own variant and development of the case. For a recent systematic theoretical defense of the possibility of KK failure, see Williamson (2000: ch. 5). For a recent defense of KK, see Greco (2014b, 2014c).

sort of disposition in test environments, and this disposition causes her to feel as if her answers are guesses.

Take the proposition that Charles I ascended to the throne in 1625 – one correct answer that Sayeqa has given in a long list of many. The verdict that KK deniers take as natural in this case is that Sayeqa knows that Charles I ascended to the throne in 1625, but that she does not know that she knows this. However, it would be wrong to say that Sayeqa's knowledge that Charles I ascended to the throne in 1625 somehow amounts to her knowing without possessing evidence that supports believing this proposition. She does possess evidence: the evidence she gained in the classroom.[83] That evidence decisively supports believing that Charles I ascended to the throne in 1625. What Sayeqa lacks is decisive evidence *that* she has this evidence – just as she lacks knowledge that she knows. But in the absence of decisive evidence that she has this evidence, and in the presence of the (misleading) impression that her answer is just a guess, it seems that Sayeqa's evidence supports believing that her evidence supports suspending judgment about the proposition that Charles I ascended to the throne in 1625.[84] So, Sayeqa's evidence supports believing that Charles I ascended to the

---

[83] In this respect Sayeqa is not quite like the "clairvoyant" who is somehow reliable about some subject matter without possessing any identifiable evidence about the subject matter (other than her own direct insights of clairvoyance). Hawthorne & Srinivasan (2013: 18) attempt to deploy the clairvoyant case to argue for something like (PIF), but I take the unconfident examinee to be at least a somewhat less tendentious case than the clairvoyant for these purposes, precisely because she does possess identifiable evidence for her first-order belief. What she lacks is decisive evidence *that* she has this evidence.

[84] This assumes that one's evidence can support believing a falsehood. I take this to be the intuitive view – misleading evidence makes possible false but justified (in the sense of evidentially supported) beliefs. For example, in the less-fancy analogues of Gettier cases where the belief in question is not luckily true, the belief is still justified (in the sense of being supported by the evidence). So, if a clock reads '5:23' but is actually stopped, and I have no reason to suppose it is stopped, and the time is not (as in the Gettier case) 5:23 by luck, my belief that it is 5:23 is supported by the evidence.

A few epistemologists (Sutton 2007, Littlejohn 2012, and on some readings Williamson himself) challenge this orthodox view. I don't find the challenge persuasive. Note that the view does *not* follow from the claim that only knowledge can provide evidence. One's body of knowledge (even though that body of knowledge consists only of truths) can make falsehoods probable to a very high degree; it just can't give them probability 1. So provided one holds that one's evidence can *pro toto* support belief in a proposition without making that

throne in 1625, but also supports believing that her evidence supports suspending judgment about (rather than believing) the proposition that Charles I ascended to the throne in 1625. So, Sayeqa is in a case of iterative failure.

Of course, someone might try to poke holes in these verdicts. For example, one might say that Sayeqa's feeling that her answer is a guess is enough to defeat the evidence she gained in the classroom at the first-order level, such that Sayeqa's evidence does not support believing that Charles I ascended to the throne in 1625. Or one might say (with considerably less plausibility, in my view) that Sayeqa's evidence gained in the classroom itself makes it the case that her evidence does not support believing that her evidence supports suspending judgment about whether Charles I ascended to the throne in 1625. It's hard to give knock-down responses to these claims, for the reasons given in the introduction to this section. In general, however, I do not see why these responses should be more convincing when it comes to *evidence* than when it comes to *knowledge*. And one can make the same protestations to deny that the case is an instance of KK failure: maintain that Sayeqa's sense that she is guessing defeats her first-order *knowledge*, or that her evidence from her classroom instruction secures higher-order knowledge that she knows that Charles I ascended to the throne in 1625. So if

---

proposition probable to degree 1, the claim that only knowledge provides evidence is entirely consistent with the claim that evidence can support believing falsehoods.

However, for those who are persuaded that one's evidence cannot support falsehood: you must still allow that one's evidence sometimes supports being *highly confident* in falsehoods. So we get the possibility of a kind of iterative failure just slightly weaker than the one I am arguing for, whereby one's evidence supports believing p while also supporting being *highly confident* that one's evidence does not support p. While this combination of attitudes does not violate (ILC) as stated, I think it is plausible that it is also a kind of rational incoherence for the same sorts of reasons that violations of (ILC) are – though I do not know how to precisely specify the bounds of this incoherence. If that is right, then cases of iterative failure still throw up a conflict between inter-level coherence requirements of some kind and evidence-responsiveness, even given the claim that one's evidence never supports outright belief in falsehoods.

the case strikes us as a convincing instance of KK failure, I also think we should take it to be a convincing instance of iterative failure of evidence.[85]

Someone might protest in reply that KK fails in this case for a reason that has nothing to do with the putative iterative failure of evidence. Specifically, the suggestion might be that KK fails in this case because Sayeqa does not even *believe* that she knows that Charles I ascended to the throne in 1625. It might then be claimed that Sayeqa is in a good enough evidential position to know this, if only she believed it. So the failure of KK is purely due to Sayeqa's not believing the higher-order proposition, and suggests no iterative failure of evidence.[86] Applied to all similar cases (as it would have to be to block (PIF)), the strategy would be to say that KK only *ever* fails for the (in one sense shallow) reason that belief sometimes fails to iterate.

I don't think this objection is right even in the case at hand. Given that Sayeqa can't *identify* that she is basing her belief on the evidence she got in the classroom (as opposed to random guessing), it does not seem that she is in a good evidential position to know that she knows that Charles I ascended to the throne. But even if one does find the objection plausible, we can complicate the case slightly so as to generate a plausible iterative failure of evidence

[85] In response, someone might claim that although they accept that Sayeqa's case is one of KK failure, when it comes to evidence it is best diagnosed as a case where, although Sayeqa's evidence supports believing that Charles I ascended to the throne in 1625, her evidence supports *suspending judgment* about whether her evidence supports believing that Charles I ascended to the throne in 1625. This does not give us the stronger kind of iterative failure I need (whereby her evidence supports believing that her evidence supports suspending judgment about the proposition that Charles I ascended to the throne in 1625). But it seems to me that if we develop the case right – so that Sayeqa really feels that she is guessing, as opposed to being *unsure* as to whether she is guessing or recalling correctly – then the latter description is really the correct one. As long as we hold fixed that in fact Sayeqa is *not* – contrary to the phenomenology she experiences – guessing, then it seems like either case can still be one in which Sayeqa's first-order belief is in fact supported by the evidence.

[86] Radford (1966), who first introduced the case, thought that Sayeqa does not even believe the *first-order* proposition in question, and that it was a case of knowledge without belief. If that's right, the KK failure cannot be due to a failure of belief to iterate, since belief is present at neither level.

even from an instance of KK failure that is only based on a failure of the iteration of belief. This exploits the role that knowledge plays in *providing* evidence.[87] Here the idea is that since one's knowledge plays a role in providing evidence for one's other beliefs, failures to know that one knows can lead to failures to know what one's evidence is, and so to iterative failure.

For this case, consider now the proposition that James I ceased to be King in 1625. Sayeqa knows, let's suppose, that James I immediately preceded Charles I, so that the year in which Charles I ascended to the throne would have to be the year in which James I ceased to be King. Given that Sayeqa knows that Charles I ascended to the throne in 1625, then, it seems right to say that Sayeqa's evidence supports believing that James I ceased to be King in 1625. However, the proposition that Charles I ascended to the throne in 1625 does not itself bear on the higher-order question of what doxastic attitude *Sayeqa's evidence* supports with respect to the proposition that James I ceased to be King in 1625. What would support believing the higher-order proposition that Sayeqa's evidence supports believing that James I ceased to be King in 1625 would be the corresponding higher-order proposition that Sayeqa's evidence includes the (first-order) proposition that Charles I ascended to the throne in 1625. But, since Sayeqa doesn't know that she knows that Charles I ascended to the throne in 1625, she is (plausibly) correspondingly unaware that this (first-order) proposition is part of her evidence. So the higher-order proposition isn't part of her evidence. So she lacks any significant evidence that her evidence supports believing that James I ceased to be King in 1625. So her evidence supports suspending judgment about whether her evidence supports believing that James I

---

[87] That knowledge plays such a role follows from Williamson's (2000: ch. 9) view that your evidence just is your knowledge, but also from various less bold claims. See Brown (2015) for a view of evidence that is critical of Williamson's whilst maintaining the core claim that your knowledge is generally part of your evidence.

ceased to be King in 1625, even though her evidence supports believing that James I ceased to be King in 1625. So, again, we have a case of iterative failure.

I hope to have just shown that (PIF) comes along very naturally with the failure of KK, as well as to have made both kinds of iterative failure fairly plausible. As I said in the introduction to this section, since the precise amount of evidence required to decisively support an attitude is hard to specify, one may quibble over the details of individual cases. So, having argued for a link between KK failure and (PIF), let me rehearse one more theoretical reason (inspired by Williamson (2000)) to expect KK failure to often occur, independently of case intuitions. The argument is premised on the plausible thought that knowledge must be "safe": to know p, it must be that one could not have easily (that is, in close possible worlds) gone wrong about whether p. The following diagram will help to illustrate why such a condition on knowledge gives us reason to expect KK failure:

line of safety for w_1
line of safety for w_A
S-knows-p-worlds
p-worlds
possible worlds
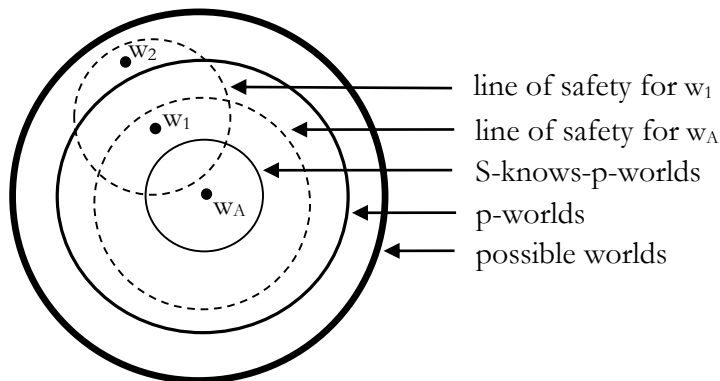
Figure 1: safety and KK failure

The p-worlds are all the worlds where p is true. The S-knows-p worlds are all the worlds in which S knows p. w_A is the actual world, and distance from w_A represents the closeness of various possible worlds. The dotted lines ("lines of safety") represent the boundary of the sphere of worlds within which S must have the correct belief about some proposition in order to know that proposition at a world. To know p at w_A, S must have the correct belief about p

78

in every world within its line of safety; to know p at w_1, S must have the correct belief about p in every world within its line of safety.

Suppose now that w_1 is a world where S believes p, and where S believes that she knows p; suppose also that w_2 is a world where S believes p. As you can see from the diagram, S's belief in p at w_1 is true (since w_1 is within the p-worlds), but her belief that she knows p is false (since w_1 is not within the S-knows-that-p worlds). The falsity of the second belief is explain by the fact that S's belief in p is unsafe at w_1: at a relevantly nearby world (w_2), S falsely believes p. As such, S does not know p at w_1. So far we have not stipulated anything that a denier of KK could claim to be impossible: all we have suggested in that subjects can have true beliefs that are not knowledge, and that they can mistakenly believe themselves to have knowledge in some such cases.

Now for w_A. As you can see, S knows p at w_A. That is compatible with a safety condition on knowledge, which requires S to be right about p in all the worlds within the line of safety for w_A. For w_1 is a world where S is right about p, whereas w_2 is too faraway to matter for whether S knows p at w_A. Nevertheless, S's belief at w_A that she knows p is unsafe, because there is a world within the line of safety – w_1 – where she falsely believes that she knows p. So S does not know that she knows p. So w_A is a world where S knows p without knowing that she knows p, generating a failure of KK.

(b) Cases of being misled about the evidential support relations

The cases we have considered so far are ones in which the source of one's ignorance about what one's evidence supports is ignorance about what one's evidence *is*. But even if one is entirely correct about what one's evidence is, one can still be misled about which doxastic

attitudes this body of evidence supports, by being misled about the *support relations* that hold between items of evidence and doxastic attitudes.

As I said in section 0.6, I take the facts about the evidential support relations to be *normative* facts. In general, it seems that one can often have misleading evidence about normative facts. A paradigm instance is when one receives misleading testimony. So in looking for a case of iterative failure, it is natural to look for a case where one receives misleading testimony about what one's evidence supports. Consider, then, the following case (loosely based on cases in Coates (2012) and Horowitz (2014)):

> **Miss Marple and Mabel.** Miss Marple is a detective who is famously good at assessing evidence. Miss Marple is investigating a murder that took place at the mansion on the hill, and she takes her great niece Mabel along with her. Miss Marple and Mabel set about the mansion collecting clues. Unfortunately, in their initial sweep of the house, nothing that they learn offers any kind of significant support to any particular hypothesis about who committed the crime. However, Miss Marple makes an uncharacteristic error, and declares to Mabel, "the clues lying around this house that you have seen today support believing that the vicar did it."

What should we say about Mabel's position? By stipulation, the case is one where the clues lying around the house do not support believing that the vicar did it. However, Mabel has excellent reason to believe that what Miss Marple says about what the evidence supports is true. So it seems that this is a case where the clues lying around the house support suspending judgment about whether the vicar did it, but Mabel's evidence supports believing that the clues

support believing that the vicar did it. The clues around the house are the crucial putative evidence that the vicar did it. And if the clues don't actually support believing that the vicar did it, Mabel's total evidence supports suspending judgment about whether the vicar did it. But Miss Marple's testimony is sufficient evidence to believe that the clues do support believing that the vicar did it, so Mabel's total evidence supports believing that her total evidence supports believing that the vicar it. That is a case of iterative failure.

Here, as with the examples in the last section, one can try to resist the case in a piecemeal way, exploiting the vagueness of how much evidential support is decisive. One can claim that Miss Marple's testimony is insufficient to support the higher-order belief that the evidence supports believing that the vicar did it. Or, going in the completely opposite direction, one can claim that Miss Marple's testimony is sufficient not just for this higher-order belief, but also for the *first-order* belief that the vicar did it. I'll begin, then, by offering some more theoretical reasons to think that this way of resisting the possibility of iterative failure won't work. This suggests that there will be a case structurally like that of Miss Marple and Mabel that amounts to iterative failure. The case of Miss Marple and Mabel is itself an illustration of the relevant structure. Unfortunately, it cannot be turned into a knockdown case simply by offering further details and precisification, due to the ineliminable vagueness of how much evidential support is decisive. But I hope to show that we have strong reasons to expect at least some cases of iterative failure that exploit the same structure. Then, I will consider some more general, principled arguments against such a possibility, and argue against them.

Let me start by giving what seems to me the right diagnosis of cases like that of Miss Marple and Mabel. It would be an overstatement, I agree, to say that Miss Marple's testimony about what the evidence supports *only* bears on the higher-order question of what the evidence supports. When Miss Marple testifies that the evidence supports believing that the vicar did it,

she gives Mabel *some* pro tanto evidence in favor of the first-order claim that the vicar did it. In this way, higher-order evidence (at least usually) "trickles down" to support first-order attitudes.[88]

However, I hold that the evidence provided by Miss Marple's testimony about what the evidence supports bears *less strongly* on attitudes towards the first-order question of whether the vicar did it than it does on attitudes towards the higher-order question of what the evidence supports.[89] After all, if what Miss Marple says is true, then that just *settles* that the evidence supports believing that the vicar did it – there is no evidential "gap" between the two. However, she could be speaking truly and yet it still be false that the vicar did it. Her testimony bears on that question without (even if true) settling it.

By contrast, the clues themselves, if anything, bear more strongly on attitudes towards the first-order question of whether the vicar did it than they do on attitudes towards the higher-order question of what the evidence supports. Primarily, the clues speak to what attitude to take toward the first-order question of whether the vicar did it (in this case, they support suspending judgment). To the extent that the evidential import of the clues is (to some degree) self-evident, they also may be able to support beliefs about their own nature, like the belief that they themselves support suspending judgment about whether the vicar did it. But this role seems secondary, and to require a greater strength of evidence to be decisive. In any event, the evidential weight of the clues at the higher-order level certainly isn't *stronger* than its weight at the first-order level.

---

[88] It doesn't always, though, at least not when we are concerned with *pro tanto* evidence: see Fitelson (2012). For some more general doubts about whether higher-order evidence trickles down in a systematic way, see Lasonen-Aarnio (2014).

[89] This moderate stance on higher-order evidence is also taken, as I read them, by Kelly (2010) and Pryor (2013: 99-100).

Putting these two points together, we get the result that, in some case at the margins, we should expect the clues themselves to outweigh Miss Marple's testimony at the first-order level, but Miss Marple's testimony to outweigh the clues at the higher-order level. That creates the possibility of iterative failure. Quibbling with individual cases by exploiting the vagueness of evidential support doesn't mitigate this possibility. However, there may be more systematic rationales for denying the possibility of iterative failure. I will now consider two such putative rationales.

The first strategy aims to establish that Miss Marple's testimony is not decisive even at the level of the higher-order attitude. The idea here is to claim that the justification of normative claims about evidential support relations is *a priori* and indefeasible by empirical testimony. Consequently, there are no false but evidentially supported beliefs about such claims. If that is right, no amount of misleading testimony from Miss Marple can outweigh Mabel's *a priori* justification for believing that the clues (given a fix on the empirical question of what the clues are) support suspending judgment.[90] This testimony is effectively inert, in the sense that it is said to be in principle incapable of tipping the balance in favor of believing that the clues support believing the vicar is guilty. What the evidence on balance supports with respect to what the clues support is fixed by one's *a priori* justification only; Miss Marple's testimony is powerless to ever change what the evidence on balance supports with respect to this matter.

---

[90] This traditionally unorthodox view that normative beliefs cannot be justified but false has recently received several defenses: from Titelbaum (2015b), Whiting & Way (ms.) and Kiesewetter (ms.-b). Titelbaum restricts his claim of infallibility to beliefs about *rationality*, whereas the other two papers defend it with respect to all normative beliefs. If one distinguished rationality from evidence-responsiveness, one might be able to accept Titelbaum's claim (for which he has a powerful argument) without committing oneself to the impossibility of justified false beliefs about evidential support. Titelbaum himself seems to assume that his view does have this commitment, though he has elsewhere defended a very subjective notion of evidential support (Titelbaum 2010), which may mitigate the radical-soundingness of this claim.

This, I think, is a very implausible verdict about the case. Mabel has excellent evidence that Miss Marple's ability to assess what the evidence supports is far superior to her own. Miss Marple is an expert – indeed, we can make her an arbitrarily reliable expert as long as we don't make her reliability *perfect* – about a particular subject matter: what the evidence supports. The idea that some person could be an expert about some subject matter to some arbitrarily reliable degree, and yet their testimony about that subject matter be in principle incapable of affecting what the evidence on balance supports, is hard to stomach.

Note also that the principle is not just about misleading testimony, but rather blocks any other kind of empirical defeat of your *a priori* justification. So, for example,[91] suppose you discover that you are on a drug that makes people form inaccurate beliefs about what the evidence supports 99% of the time. As it happens you are in the lucky 1% who aren't affected. Your belief about what the evidence supports feels justified to you, but it would feel equally well-justified if you were on the drug and it were mistaken. In a case like this, the current strategy is committed to saying that your discovery that you are on the drug does nothing to evidentially undermine your belief about what the evidence supports.

More generally, consider the sorts of truths that the evidential support relations express. Remember that the position we are considering requires not just indefeasible *a priori* justification for *general* normative claims like "believe what the evidence supports", but for the detailed normative facts of which pieces of evidence support which attitudes and to what degree. Such truths are, in my view, often radically unobvious to us. For example, what is the evidential support relation between the proposition that it seems to you that you have hands

---

[91] This well-known sort of case originates from Christensen (2007: 10).

and the proposition that you have hands?[92] What is the evidential support relation between the proposition that a theory is simple and the proposition that the theory is true? What is the evidential support relation between the proposition that past scientific theories failed and the proposition that our current theories are false? What is the evidential support relation between evidence that is very selectively filtered in biased ways and one's beliefs?

I think we are in pretty bad epistemic positions with respect to these questions – much worse epistemic positions than we face with respect to common empirical questions. Many people would have no idea how to answer them whatsoever. As such, the best evidence they have to go on about them will often be testimonial and by upbringing – by learning from others what to take as evidence for what. So I think it is highly implausible to say that these questions are ones about which everyone possesses indefeasible *a priori* evidence that renders (potentially misleading) empirical evidence irrelevant.

Moreover, even if one bites the bullet on these claims, this move will, I think, simply push the conflict between coherence and evidence-responsiveness elsewhere. Consider again the case of Mabel. Suppose that Mabel herself (correctly) judges Miss Marple to be much more reliable than she herself is in judging what the evidence supports. This itself carries certain commitments as to how to adjust her beliefs about what the evidence supports when she finds them to conflict with Miss Marple's. So, if she fails to treat Miss Marple's testimony as significant evidence with respect to the matter of what the evidence supports, without greatly downgrading Miss Marple's expert status (which is surely not warranted by one disagreement), she will be guilty of a different kind of incoherence.[93] So, if one claims that the evidence

---

[92] This, of course, is not a marginal example brought on only by thinking about brains in vats: it generalizes to every single case of using appearances as evidence.

[93] See Worsnip (2014) and the conclusion of this dissertation, part b.

supports Mabel maintaining her belief that the evidence supports suspending judgment in this case, even in the face of Miss Marple's testimony to the contrary, one purchases concord between evidence-responsiveness and (ILC) only at the price of conflict between evidence-responsiveness and other coherence requirements. So the ultimate claim I am making that evidence-responsiveness and coherence requirements conflict remains plausible even in the face of this strategy.

So much for the first strategy. The second strategy, moving in the complete opposite direction, attempts to provide some general principle that establishes that Miss Marple's testimony will automatically defeat the justification provided by the clues themselves even at the first-order level. What could such a principle be? We might look here to a principle suggested by Bergmann (2005), that whenever one has a higher-order belief that one's first-order belief is unjustified, that is sufficient to defeat one's justification for the first-order belief. (Bergmann seems to be treating justification as equivalent to evidential support here.) Since Miss Marple's testimony produces a belief in Mabel to the effect that her first-order attitude is unjustified, the claim would then be that this belief itself defeats her justification for her first-order attitude.

However, as you may have noticed, there is a complication here. Bergmann's original claim pertains to the defeat of first-order *beliefs*. But in our example, the first-order attitude supported by the clues is not *belief*, but *suspending judgment*. The clues support suspending judgment about whether the vicar did it, but Miss Marple offers misleading testimony that they support believing that the vicar did it. One might think this is not a problem: just extend Bergmann's claim to all first-order attitudes rather than just beliefs. But this has a nasty result. If Mabel believes that her evidence supports believing the vicar did it, the extended version of Bergmann's claim entails that this belief defeats any supports she had for suspending

judgment. It would do the same for any putative evidence for active disbelief that the vicar did it. But, assuming that the evidence always supports at least one attitude (and I explained in section 1.2 why this is a consequence of the way I am talking of evidential support), that means that Mabel's belief that her evidence supports believing that the vicar did it guarantees that the evidence does support believing that the vicar did it. In other words, Mabel's belief makes itself true, and is infallible.

Bergmann himself disavows this result (2005: 426), but the only way he can avoid it is by maintaining his restriction of his principle to beliefs rather than all attitudes including suspension of judgment.[94] For whatever reason, it seems more plausible to us that a *belief* can be "defeated" by mere presence of conflicting higher-order beliefs (so that one's evidence supports suspending judgment) than it does that a *suspension of judgment* can be "defeated" by mere presence of conflicting higher-order beliefs (so that one's evidence support believing). This is perhaps because "defeat" naturally sounds like something that takes away support for something *positive*, and leaves us with the more "negative", non-committal attitude of suspending judgment – rather than something that takes away support for that non-committal attitude and leaves us with the more committal attitude of belief. However, blocking (PIF) would require a principle that promises higher-order defeat of all doxastic attitudes, not just belief. In the case of Miss Marple and Mabel, I deliberately chose an example where the first-order evidence supports suspension of judgment rather than belief, to make the higher-order defeat strategy harder to execute. Now, it may be that the higher-order defeat strategy also

---

[94] Cf. also Huemer (2011: 7). Huemer appeals to the principle that "if one has good reason to doubt that one's belief that P constitutes knowledge, then one thereby has a defeater for one's belief that P." It's hard to see how one could adapt this principle for suspension of judgment, since a suspension of judgment cannot constitute knowledge. Surely the idea could not be that when I have any good reason to doubt that I should be suspending judgment, suspending judgment is "defeated", so that the evidence suddenly supports outright belief (or disbelief).

fails to block (PIF) when the first-order evidence supports belief (indeed, that is my view). But even if it only fails when the first-order evidence supports suspending judgment, that is enough to leave (PIF) unimpeached. Such cases will still be iterative failures, and that is enough to get the conflict between evidence-responsiveness and coherence going.

Bergmann is right to think that there is something "bad" about taking an attitude that you believe your evidence not to support. But instead of trying to cash this badness out as a principle whereby the latter, higher-order belief automatically defeats any evidential support for the former, first-order attitude, we should capture the badness through a *sui generis* coherence requirement, (ILC). Then we can recognize the distinctive badness of this *combination* of attitudes, rather than implausibly taking the higher-order attitude to be authoritative or infallible with respect to the *reasons* for the first-order attitude.

That brings me to the end of my defense of (PIF). I have argued that iterative failure can occur for two distinct reasons: either because one is misled about what one's evidence is (as in cases of KK failure such as that of Sayeqa, the unconfident examinee), or because one is misled about the evidential support relations (as in cases of misleading testimony such as that of Miss Marple and Mabel). If either contention is right, that is sufficient to establish (PIF).

1.5   *Does (ILC) still have force in cases of iterative failure?*

As I mentioned in section 1.3, my argument would be in trouble if the rationale for (ILC) broke down whenever iterative failure obtained. So to complete my defense of (ILC), I want to briefly illustrate the way that it does not do so.

The crucial point here is that, at least in most cases,[95] one cannot know that one is in a case of iterative failure. Let's illustrate this with respect to Sayeqa and Mabel. In Sayeqa's case, iterative failure is generated by a failure of KK. The proof that one cannot know that one is in a case of KK failure is simple. Let p be the proposition that KK fails with respect to (in Sayeqa's case, that Charles I ascended to the throne in 1625), and let $K_S(p)$ be the claim that Sayeqa knows p. By the definition of KK failure:

(1) $K_S(p)$

(2) $\neg K_S(K_S(p))$

Suppose for reductio that Sayeqa knows that she is in a case of KK failure with respect to p, Sayeqa would have to know (1) and (2). In other words:

(3) $K_S(K_S(p))$

(4) $K_S(\neg K_S(K_S(p)))$

But (2) and (3) contradict each other. So, by reductio, Sayeqa cannot know that she is in a case of KK failure with respect to p.

Similar considerations apply to Mabel. For Mabel to know that she is in a case of iterative failure, she would have to know both that her evidence supports suspending judgment about whether the vicar did it and that her evidence supports believing that her evidence

---

[95] One could argue that a case of iterative failure given by Williamson (2011: 153-4) constitutes an exception to this. Horowitz (2014) carefully points out some important differences between this case and more standard cases of putative iterative failure. I refrained from relying on this precise kind of case in my arguments for (ILC) above.

supports believing that the vicar did it. But if Mabel knew that her evidence actually supports suspending judgment about whether the vicar did it, she would be in a position to discount the misleading evidence that supports believing that her evidence supports believing that the vicar did it – in which case it wouldn't be misleading evidence at all, and she wouldn't be in a case of iterative failure.

The general lesson here is that iterative failure depends crucially on being misled in a way that precludes recognizing that you are in such a situation. This is significant. We might have a case for the failure of (ILC) if cases of iterative failure allowed one to provide a rationalization of why one is violating (ILC). We might imagine Mabel saying, "I see that normally I shouldn't violate (ILC), but this is a strange case where my evidence supports a particular doxastic attitude, but my evidence also supports believing that my evidence does not support this attitude." But actually, Mabel is not in a position to say this, for the reasons given above.

When one violates (ILC), one by definition takes oneself to be violating (ER). Imagine, then, what it would take to actually violate (ILC) (and respect (ER)) in a case of iterative failure. According to (ER), Mabel is rationally required (i) to suspend judgment about whether the vicar did it and (ii) to believe that her evidence supports believing that the vicar did it. Suppose that Mabel actually does this, as (ER) demands of her, and that she is aware of her own doxastic states. In that case, Mabel *takes* herself to be violating (ER): as far as she is concerned, this is just any old case where she's failed to believe what her evidence supports. But since she is in a case of iterative failure, it turns out that by luck she does satisfy it, in spite of herself.

It's an interesting question whether one can ever be in the position that (ER) diagnoses this case as putting Mabel in – namely, where the only way to be rational is by luck, in spite of

90

believing oneself to be failing.[96] If one cannot be in such a position, this presents a further reason to suspect that (ER) is not best understood as a requirement of *rationality*. Regardless of how we answer this question, however, the point I want to make is that this kind of lucky satisfaction of (ER) certainly does not seem to diminish the intuitive irrationality of her violating (ILC). Mabel can give no rationalization for her violation of (ILC) in terms of her need to satisfy (ER), since by her lights she thinks she is violating (ER). So the incoherence of her (ILC)-violating mental states is just as clear in this case as it is with any other violation of (ILC). And the argument for (ILC) given earlier in terms of the relationship of a judgment of evidence to a judgment of likelihood is unaffected.

1.6   *What kind of conflict?*

Let's take stock. I've argued that cases of iterative failure are possible, but that (ILC) is nevertheless a genuine coherence requirement. This shows that evidence-responsiveness (responsiveness to epistemic or evidential reasons) and conformity to coherence requirements can come into conflict with one another: there are cases where one cannot have both. So the view sketched at the start, on which evidence-responsiveness guarantees coherence, is false. So we cannot say that rationality consists solely in evidence-responsiveness, and expect that to capture our intuitions about the irrationality of certain forms of incoherence derivatively. That already provides us with decisive reason to reject Evidentialism, the view that one is rational iff one satisfies (ER). It also shows that coherence requirements will have to be theorized, at least to some degree, in their own right.

---

[96] I do genuinely mean that it is interesting. I am not coyly expressing contempt for the view that such a case is possible. For some considerations that may support believing that it is, see Hawthorne & Srinivasan (2013); Srinivasan (forthcoming).

As I mentioned in section 1.1, there are two ways of understanding this conflict between coherence requirements and evidence-responsiveness. One is to say that both coherence requirements like (ILC) on one hand, and (ER) on the other, are genuine rational requirements, in some recognizable and unified sense of the term 'rational'. Then, cases of iterative failure will count as rational dilemmas – cases where whatever one does, one is not rational. The second possibility is to say that really there are two fundamentally different kinds of normative demand here, that need to be pulled apart. In that case, we have conflicts *between* two different normative domains, but not *within* one single normative domain. The first diagnosis is, on one reading, endorsed by David Christensen (2007, 2010, and esp. 2013: 92-96). Christensen's view is that both (ER) and (ILC) are "rational ideals", and that in cases of iterative failure, one "will end up violating some ideal or other." I will now argue for the second diagnosis over the first.

The first argument in favor of the second approach over the first is simple: it avoids our having to posit rational dilemmas, whereby an agent is irrational *whatever* she does. This might seem like hair-splitting, since on the second approach we still have to acknowledge conflicts *between* different normative domains, such that agents must fall short of normative ideality in *some* respect. I'm not sure that this is as bad a thing to have to say: there does seem to be something about *irrationality* in particular that should be in some way due to the agent herself, and not simply due to a situation outside her control. In any case, though, there are formal advantages to representing normative conflicts as occurring between different normative domains (as "inter-domain conflicts"), rather than within a single domain (as "intra-domain conflicts). For example, it allows us to preserve the axioms of standard deontic logic. If you are rationally required to Φ and rationally required not to Φ, then by standard deontic logic you are rationally required (to Φ and not to Φ). In other words, you are rationally required

to do the impossible. If, conversely, the normative requirement to Φ is of a fundamentally different kind to the normative requirement not to Φ, then there is no *single* sense in which you are required, even by the lights of standard deontic logic, (to Φ and not to Φ).

Furthermore, allowing *just any* kind of conflict between requirements is permissive in a way that makes the methodology of arguing for particular requirements considerably more difficult. One check on our ability to posit rational requirements costlessly is the possibility that such rational requirements might conflict with other, more plausible, requirements. If we allow for rational dilemmas, then we can never show a putative requirement of rationality to be false by showing that it conflicts with some other important requirement. By contrast, if we allow conflicts across normative domains but not within them, the ban on intra-domain conflicts will still give us some way to rule at least some putative requirements out.

The second argument in favor of the second approach over the first is that coherence requirements and evidential reasons are simply metaphysically distinct in a way that the first approach is likely to obscure. I have said a bit about this already in section 0.3, and will be saying more in chapter 2. As I'll argue there, the charge of incoherence carries a different sort of criticism than the charge of having failed to respond correctly to one's evidence does. The metaphysical distinctness of these two normative kinds is also reinforced by the scope difference that I called attention to in sections 0.3 and 1.3.

Relatedly, as I'll return to in section 2.1 below, coherence requirements and evidential reasons play very different sorts of roles in how they guide reasoning. As Niko Kolodny (2005: see esp. 547) has argued, one is not usually guided in deliberation about what to believe by the thought that by believing something, one can avoid incoherence of the sort banned by (ILC). Rather, one is guided by the thought that there is weighty evidence that the proposition to be believed is true. Nevertheless, this process is in one sense *regulated* by one's background

disposition to satisfy (ILC) – that is, to take the doxastic attitudes that one *takes oneself* to have conclusive evidence for, and to refrain from taking those attitudes that one takes oneself not to have conclusive evidence for. This difference in the role that coherence requirements and evidential reasons play in reasoning reinforces their metaphysical distinctness. Though I have argued that they can *conflict* in the sense that sometimes one cannot satisfy both, they do not *compete* at the same level of one's deliberations.

This point is reinforced further by the related fact that when one satisfies (ILC), it will *seem* to one that one is satisfying (ER), and when one violates (ILC), it will seem to one that one is violating (ER). Consequently, as I already argued in section 1.5, it is normally not possible for one to know that one is in a situation of iterative failure, such that the two conflict. These considerations all taken together make it, in my view, implausible to think of (ER) and (ILC) as, in Christensen's language, "competing ideals" of the same kind which one has to somehow weigh against each other in deliberation.[97]

Last, the puzzle that we have been considering throughout illustrates how confusing coherence requirements with claims about reasons is liable to lead to substantive philosophical mistakes. Some have thought that (ILC) can be simply be restated as a claim about reasons:

Inter-level reasons (ILR)

---

[97] At times Christensen writes as if it is a *further* question, given a conflict of ideals, which ideal it would be rational to satisfy (Christensen 2013: 96). Put this way, it seems that Christensen might best be interpreted not as affirming the possibility of rational dilemmas after all (hence my slightly equivocal approach to describing him as a defender of this view). Rather, his view could actually be understood as rejecting both (ER) and (ILC) – as they are stated, as stringent *requirements*. Instead, each is an ideal – where what this means is that satisfying it is an epistemic good that one has *pro tanto* reason to realize. Indeed, Christensen (2007: 24) appeals to analogies that are clearly mere *pro tanto* conflicts between reasons to try to take the sting out of his view. Obviously, *pro tanto* reasons conflict all the time, without giving rise to deontic "dilemmas" in the strong sense of the term. However, for the reasons I have just given in the text, this seems an implausible account of the relationship between (ER) and (ILC): they do not seem to "trade off" against each other in deliberation in the right way.

(i)   If S has most epistemic reason to believe that her evidence supports

        D(p), then S has most epistemic reason to take D(p).

(ii)   If S has most epistemic reason to believe that her evidence does not

        support D(p), then S has most epistemic reason not to take D(p).[98]

Given the equivalence of what one's evidence (*pro toto*) supports believing to what one has most epistemic reason to believe, the second part of (ILR) just is the denial of (PIF). So if one takes the inter-level principle to statable as a claim about reasons, it is clear why one would think that the inter-level principle is incompatible with (PIF), such that one has to choose between them, and such that the acceptance of one provides conclusive grounds for the rejection of the other. But although either (ILC) or (ILR) might be picked out by the vague term "level-bridging principles", the two claims are really of fundamentally different kinds. Whereas (ILC) is a wide-scope coherence requirement, (ILR) is not, strictly speaking, a requirement at all, but rather a strong metaphysical constraint on the reasons that it is possible for an agent to have at any given point in time. Specifically, (ILR) posits the inability to be evidentially misled about your own epistemic reasons in a systematic way. This is simply not a commitment of (ILC).

    If a philosopher confuses (ILC) with (ILR), and moves between them uncarefully, she will be liable to be lead to the substantively mistaken philosophical view that (ILC) and (PIF) are incompatible claims. This is what has, I think, lead so many philosophers to argue from one to the rejection of the other. They have felt forced into the false dichotomy of either putting an implausible constraint on their theory of evidential support (by saying that iterative

---

[98] See, e.g., White (2007: 120). Dan Greco also suggested this to me in correspondence.

failure is impossible), or into saying that there is nothing irrational about holding doxastic attitudes in the face of one's own judgment that one has no good (evidential) reason for having such attitudes. Both of these, I have argued, are false substantive views. The importance of avoiding this mistake and others like it provides us with another good reason to sharply distinguish conformity to coherence requirements from evidence-responsiveness, and not to place them under a single undifferentiated heading of 'rationality'.

1.7   *The disambiguating response revisited*

In section 0.2, I mentioned the *disambiguating response* to the coherentist view of rationality, which charges that the issue is just terminological. The thought is that one can use 'rationality' to refer only to coherence, thereby excluding (ER), but that this is just a particularly narrow usage of the term.

I have not disagreed that there are different ways of using the term 'rationality', nor that how to use the term is at least to some degree a matter of stipulation (though I will argue in section 2.3 that coherence is actually more central to our ordinary concept of rationality and our ordinary usage of 'rational'). However, I hope to have shown that on its own, the disambiguating response passes over much that is of substantive importance. First, the presupposition that the coherentist notion of rationality is just some *narrow* or *weak* sense of the corresponding term is mistaken in that coherence is not guaranteed by reasons-responsiveness. Second, a notion of rationality that is supposed to cover both reasons-responsiveness *and* coherence is not available, at least not if we want to avoid positing rational dilemmas and avoid lumping metaphysically distinct phenomena together. Indeed, if one rejects the possibility of rational dilemmas (or of intra-domain deontic conflicts more broadly),

(ER) and (ILC) are simply not compatible as putative requirements of rationality – or so I have argued.

The arguments for these points did not *begin* by assuming or stipulating that rationality consists only in obeying coherence requirements. They simply began by assuming that coherence requirements – and specifically, (ILC) – are at least one thing that we want our theory of rationality to capture; an assumption shared by many of my opponents. If that is right, and I am right that we should not use 'rationality' to capture both rationality and evidence-responsiveness, and we have a different normative vocabulary – that of reasons – for capturing the significance of evidence-responsiveness, then I think we should use 'rationality' to refer to the satisfaction of coherence requirements, so that we do not lose something important and distinctive.

It is not unusual for philosophers to make clarifying stipulations about terminology relative to more inchoate ordinary usage. Consider the ordinary English word 'logical'. People do sometimes use the term 'logical' to mean at least approximately what philosophers mean by it, but other times it seems that broadly speaking it is being used to refer to being rational or being sensible. People talk of 'the logical thing to do', or of a policy being logical, neither of which make any sense at all on the philosophical meaning. Yet no-one thinks that this is a deep problem for the discipline of logic, or that it need constrain our logical theories. Philosophers have been able to isolate a genuinely unified subject-matter in the discipline of logic, a subject-matter that is profitably distinguished from the study of being sensible. Consequently, we don't take it to be a mark against a logical system that it fails to count everything ordinary people describe as 'illogical' as a violation of a law of logic.

That doesn't mean that we have to wage some campaign to root out any ordinary usage of 'logical' (or, indeed, of 'valid') that doesn't give it its philosophical sense. Nevertheless, since

logic in the philosophical sense is a phenomenon of distinctive interest, we want to be a sure to have a name for it in philosophy and not to run it together with the study of being rational or sensible. If we did that, we would lose something of great theoretical and explanatory interest. I suggest that the case of rationality – and the need to distinguish it from responding to reasons and evidence – is analogous.

Some might now be inclined to respond, instead, by suggesting that unlike logic, coherence requirements are simply not of much epistemological or broader philosophical interest, and should be sidelined from the theory of rational belief and of doxastic normativity more generally. I'll be trying to say more about why coherence requirements *are* of philosophical interest in the next chapter, and throughout the remainder of the dissertation as a whole. However, I do hope to have already shown at least one instance in which coherence requirements are epistemologically interesting, by showing how conceiving of (ILC) as a coherence requirement, and not as a claim about reasons, and distinguishing the two clearly, can help us to resolve the puzzle about higher-order evidence that we began with in section 1.1. So let's conclude this chapter by recapping what the solution to that puzzle, in the end, is.

The puzzle, recall, was as follows. It seems that cases of "iterative failure," in which one has all-things-considered misleading higher-order evidence, are possible (PIF). That is: there can be cases where one's evidence supports some doxastic attitude D toward some proposition p, but one's evidence supports believing that one's evidence does not support D toward p. Yet, if this is possible, and rationality requires believing what one's evidence supports (as suggested by the requirement ER), then it seems that rationality requires one to take D toward p, but also to believe that one's evidence does not support D toward p. But that seems to be an irrational combination of doxastic attitudes to hold (as suggested by the requirement ILC).

On the view I have ultimately argued for, this puzzle is ultimately to be (dis)solved by rejecting (ER) in its original formulation, and replacing it with (ER*) – which formulates the central normative claim about evidence-responsiveness as one about reasons, not rationality. In cases of iterative failure, we should say, there is simply a conflict between responding to one's reasons, and being rational.

Some may find that unsatisfying: "but what *ought* I to do in such a case?," they will ask. If this 'ought' is the 'ought' of "having most reason", the answer is clear: you ought to believe what your evidence supports. Unfortunately, the tragedy of having all-things-misleading higher-order evidence is that you cannot do this (at every level of your beliefs) *rationally*.[99] That is what this kind of epistemic ignorance about what your own evidence supports consigns you to. That's the bad news.

The good news, though, is that this kind of agonizing conflict won't be tearing you apart any time soon. For, as I noted in section 1.5 above, it is also part of the misleadingness of misleading higher-order evidence that you cannot know it is misleading. As such, when conflicts between satisfying (ILC) and responding to your evidence arise, you will not know that they have arisen. If you are rational, you will go on satisfying (ILC), and in virtue of that, thinking that you have responded to your evidence correctly. From your perspective, there will be no deliberative dilemma about whether to satisfy (ILC) or whether to take the attitudes that your evidence supports. That's the good news. But you will be wrong: unbeknownst to

---

[99] What about what one *ought* to believe? As I mentioned in section 0.6, and will explain in chapter 3, I think that 'ought' is context-sensitive, in a way that makes it sensitive to information. So there will be one usage of 'ought to believe' that is relative to your evidence, but it is only one usage, and may not be privileged in any particular way. I think we can also, amongst other things, use 'ought' in a way that makes coherence requirements fundamental; much more on this chapter 3.

you, you will not taken (all) of the attitudes that your evidence supports. That's the bad news…you get the idea.

Of course, once one decides to construe rationality in terms of conformity to coherence requirements only, it is fairly obvious that (ER) has to be replaced with (ER*), as no doubt many epistemologists would recognize.[100] And the idea that different normative concepts feature in the two requirements may not seem especially tendentious. Nevertheless, as I argued in section 1.6, in assuming that one of (ILC) and (PIF) has to go, epistemologists have been tacitly taking it for granted that the same normative concept should be used for talking about evidence-responsiveness and for stating requirements like (ILC). Otherwise there is simply no incompatibility between (ILC) and (PIF), and both can be affirmed as true. So the higher-order evidence debate is one place where substantive epistemological mistakes have resulted from a failure to recognize reasons and coherence requirements as distinct normative phenomena.

This reveals a final way in which the disambiguating response on its own is too glib as a defense of the epistemological *status quo*. Suppose that a defender of (ER) says that she is just using 'rationality' in a different way to me, and that all along she meant it to be interpreted as (ER*). The problem here is that the arguments from (ER*) and (ILC) to the rejection of (PIF), and from (ER*) and (PIF) to the rejection of (ILC), are invalid. In order for such an argument to be valid, one really does need to assume that (ER) is true on the *same* meaning of 'rationality' that features in (ILC). So even if we are "merely" disambiguating (and I'll come back to that contention persistently in the next chapter), disambiguation actually reveals the invalidity of the standard arguments in the literature. It does not leave everything as it was.

---

[100] *Pace* ultra-subjectivist or coherentist construals of evidential support (cf. section 0.4).

Chapter 2
Rationality as Coherence Defended

2.1 The interest and importance of coherence requirements

This chapter takes up the task of defending the broader idea that coherence requirements are of crucial philosophical interest and importance, and of eliciting the intuition that they are central to our basic concept of rationality. Indeed, I try to show that there is at least *a* notion of rationality worth theorizing and caring about that involves *only* conformity to coherence requirements. On the way, I say more about what coherence is and what makes coherence requirements distinctive.

The plan is as follows. In this section, I'll begin by arguing that coherence is a distinctive and philosophically important notion. In 2.2, I'll say something about the question of whether coherence requirements are "normative". In 2.3, I'll argue that coherence is central to the ordinary notion of rationality, and that this ordinary notion actively tells against the idea that failures to respond to reasons always count as irrational. I'll argue that this is especially vivid in the moral case. Though I think that it persists to a lesser degree in intuitions about the epistemic case, I accept that there is an asymmetry between our intuitions about the moral and epistemic cases. In 2.4, I'll argue that this asymmetry cannot be vindicated. In 2.5, I'll supplement this by arguing that an error-theoretic story can be told to explain why the intuitions appear less coherentist in the epistemic case than in the moral case. Finally, in 2.6, I'll survey various weakenings of the reasons-responsiveness account that are designed to provide alternative explanations of the intuitive data. I'll argue that these "modified reasons-responsiveness accounts" fail, and indeed fail in ways that themselves illustrate the advantages of the coherentist account.

So let's start by saying something more general about the interest and importance of coherence requirements. Some philosophers think that coherence alone is not of much interest, and do not see why it is, in and of itself, worth caring about or wanting. Caring about coherence *per se* can seem like a kind of fetishism for mental order over the things that really matter in life: the truth and the good.[101]

This worry can be broken down into two distinct strands. On one interpretation, the question is: what is the *normative* importance, for an agent, of being coherent? On a second interpretation, the question is: what is *philosophical* importance, for a theorist, of coherence? These questions are largely independent of each other. Many notions are of philosophical importance but not of much normative importance. The ambiguity and slipperiness of the 'we' in the question "why should we care?" as voiced by a philosopher is liable to elide this independence.

I am more committal on the philosophical importance of coherence than on its normative importance. So I will begin, in this section, by explicating some reasons to take coherence to be philosophically important, regardless of how one answers the normative question. In the course of this task, we will be able to say something more about what coherence is and its distinctiveness in contrast to reasons-responsiveness. In the next section, 2.2, I will make some more limited remarks about the question of the normative importance of coherence.[102]

---

[101] Cf., e.g. Kolodny (2008a): "If we tend, or ought, to be formally coherent, it is not as such, but only as a by-product of pursuing what ultimately matters: believing the truth and choosing the good." See also Raz (2011: ch. 8).

[102] Some of the points I'll make in this section and the next have quite a bit in common with the views of Michael Ridge (2014: ch. 8, esp. 234-240), who also defends a conception of rationality as coherence, and makes some similar points in explaining why such a notion is distinctive and interesting. I did not discover Ridge's chapter until this chapter was already written, so our arguments were developed in isolation, but I wanted to acknowledge the similarities between our views. There are also, naturally, some important differences.

To begin, note that to accuse an agent of incoherence expresses a distinctive kind of *criticism* of that agent – different from that of charging them with having failed to believe or do what they have most reason to believe or do. When a speaker accuses an agent of having failed to do what she has most reason to do, she criticizes that agent as having failed by the lights of the speaker's normative standards and cares. Of course, at least on many views, when expressing this criticism, the speaker takes these standards and cares to be authoritative. Nevertheless, if the agent disagrees with this assessment and reflectively rejects the speaker's standards, there is no reason to expect the accusation of failure to move the agent unless it causes her to revise her own standards. By contrast, an accusation of incoherence amounts to the criticism that the agent fails, in some sense, by her own lights – or independently of any particular substantive standards of any sort.

We need to be precise about what exactly this means. It might be doubted that there is a real contrast between a criticism of incoherence and a criticism of failing to do what one ought to do here. It might be conceded that if someone *accepts* that he is being incoherent, he will feel a pressure to revise his attitudes. But the same is true, it might be charged, if someone *accepts* that he is not doing what he ought to do. In either case, however, someone might reject the claim that he is really being incoherent, or failing to do what he ought to do. Just as someone might accept different claims about what he ought to do, he might accept different claims about the coherence requirements of rationality. But we must not say that the agent is only irrational if he *accepts* that he is violating a coherence requirement;[103] otherwise we will get into a regress, with the result that there are no stable coherence requirements that it is always

---

[103] Cf. Davidson (2004: 177), who seems to get dangerously close to saying this. However, see also (*ibid*.: 195), from a chapter written three years later.

rational to violate.[104] So the criticism of incoherence cannot be that the agent fails by his own lights.[105]

This line of criticism is, in my view, correct right up until the last step. Clearly, it cannot be right that just calling someone incoherent automatically gets them to revise their attitudes, and we should not say that one is incoherent only if one believes oneself to be incoherent. But this is not the only way that charges of incoherence could be distinctive in the pressure they put on someone to revise their attitudes. Rather, the claim should be that for *genuine* cases of rational incoherence – where the accusation of incoherence is actually on target – the pointing out of the *substance* of the incoherence puts pressure on the person to revise their attitudes. What is distinctive is not that the recognition of the *fact that the attitudes are incoherent* results in a pressure to revise – that is true also of recognition of failing to do what you ought to do. Rather, what is distinctive is that the mere recognition that you *have the attitudes in question* results in a pressure to revise.

---

[104] To spell this out: suppose that violations of (say) the instrumental requirement are only irrational when the agent *accepts* that such a violation is incoherent. Then the instrumental requirement is not, in fact a genuine requirement of rationality. Rather, the genuine requirement adds to the antecedent of the conditional that one must believe that violations of the instrumental requirement are incoherent. But then we can repeat the same procedure with respect to the new requirement we find ourselves with: in order for violations of *this* requirement to be irrational, we might say, one must accept that violations of *this* requirement are incoherent. But then *this* requirement is also not a requirement of rationality; again, we need to add another clause to the antecedent. And so on. This regress is exactly that urged on us by the tortoise in the fable told by Blackburn (1995), who is himself constructing a practical analogue to the fable told by Carroll (1895) for the case of inference rules such as *modus ponens*. Blackburn seems to think that this poses a problem for the instrumental requirement and for requirements of rationality more broadly. (Cf. also White's (2007: esp. 123-4) criticisms of a view like coherentism, which seem to assume that it can't have fixed stopping points of this sort.) But I think that the way to avoid the regress is simple: do not let it get started. We should simply insist that violations of the (original) instrumental requirement are irrational, irrespective of whether the agent believes such violations to be incoherent or irrational. Incoherence does not require one to believe that one is incoherent; one can be mistaken about whether one is incoherent. One can also mistakenly think that one *is* incoherent when one in fact isn't.

[105] See Broome (2013: 91-3).

So, take, for example, the requirement not to have contradictory beliefs, in the strict and narrow sense of believing p and believing not-p.[106] The idea here is that once you recognize that you believe p and believe not-p, you feel a rational pressure to revise your attitudes. You don't need to accept the theoretical proposition that this constitutes a violation of a coherence requirement. So the sense in which you fail by your own lights is not that your belief set fails by the lights of a coherence requirement that you yourself explicitly accept. Rather, it is, loosely speaking, that one of your beliefs fails by the lights of the other, and vice versa. That is the incoherence in question.

Interestingly, this suggests a kind of naturalistic analysis of what it is for a set of mental states to be incoherent: the mental states that count as jointly incoherent are those which a subject cannot reflectively and consciously acknowledge that she has, without feeling pressure to revise at least one attitude. This helps us to get a grip on what incoherence is, and suggests a possible naturalistic realist metaphysics of coherence requirements. It also gives us an epistemology for identifying whether some putative coherence requirement is in fact a genuine coherence requirement. The method is to see whether the set of mental states that the putative requirement claims are jointly irrational are in fact hard to sustain reflectively under conditions of full conscious transparency to oneself.[107]

---

[106] I endorse this requirement in section 5.9 below.

[107] This is helpful because the coherentist method of enumerating requirements of rationality can sometimes feel somewhat piecemeal. One may wonder what really unifies the diverse set of requirements that make it onto the list, and what qualifies them as genuine coherence requirements. As far as I know, the only existing theorists that give a general method for discovering coherence requirements are Easwaran & Fitelson (2015), who present the intriguing suggestion that mental states are jointly incoherent when we can know, independently of knowing what the substantive reasons in the particular case support, that someone who holds those states *together* cannot be doing (or believing) what their substantive reasons support. Unfortunately, if the arguments of chapter 1 hold, this neat method won't work. For if there are conflicts between coherence requirements and substantive reasons, there are cases where a subject can violate a coherence requirement without failing to do (or believe) what their substantive reasons support doing. As I argued, cases of iterative failure are instances of this.

Neither this metaphysics nor this epistemology obviously transfers over to substantive reasons and responding to them. Suppose that someone has most reason to comply with some moral demand, like that of avoiding eating meat (you can substitute some other demand that you think issues most people with decisive reasons if you don't agree about this one). Even if this is true, it is clear that the mere recognition that you are eating meat is not enough to guarantee that you feel a pressure to stop. You need to accept the further theoretical proposition that this constitutes something that you ought not to do. Without this, you in no sense fail by your own lights. With it, you are in violation of a coherence requirement – specifically, the enkratic requirement. So substantive reasons and coherence requirements are not on a par here. This suggests, interestingly, that the metaethics associated with these two normative phenomena may be quite different.[108]

This contrast has another important upshot which helps to explicate the distinctive importance of coherence requirements. Because coherence requirements are hard to transparently and reflectively violate, agents are disposed to conform with them under conditions of full mental transparency. This means that they are poised to play certain *explanatory* roles that substantive reasons are not. We will see an illustration of this in chapter 4, where I call on the same coherence requirement that we explored in chapter 1 – inter-level coherence – to explain the difficulty of believing at will. But it also means that coherence is of importance in a number of other settings.

For example, philosophers of mind and action often make the familiar point that our capacity to interpret other agents, and to attribute mental states to them on the basis of their

---

[108] See Ridge (2014: ch. 8), and (somewhat less obviously) Blackburn (1998: esp. ch. 3) for examples of philosophers who both offer different metaethical accounts of one phenomenon as compared with the other phenomenon.

linguistic and non-linguistic behaviour, relies upon an assumption that those agents are by and large rational.[109] If one thinks of rationality in terms of reasons-responsiveness, this can seem like a puzzling assumption. We fail to respond to our reasons utterly persistently. But, in my view, the best version of this interpretationist theory will say that it is *coherence*, not reasons-responsiveness, that needs to be assumed in order to attribute mental states to agents. This fits with the idea that we can attribute some individual mental state to an agent by backgrounding other mental states of the agent and then observing her behavior. What we are assuming here is that her mental states (some of which are manifested in behavior) fit together in the right way; that is, that they are coherent.

This explains the popular thought that irrationality ultimately amounts to a kind of unintelligibility or failure of agency; as failing to be explicable or to have your actions and mental states be predictable and attributable on the basis of a standard explanatory model. To be systematically irrational is, ultimately, to be in one way unintelligible as an agent. This is the deep, core notion of rationality – in one way more fundamental than the notion of doing what one ought to do – that coherentism draws our attention to: rationality as a precondition of genuine agency.[110] There is nothing unintelligible about simply failing to respond to your reasons;[111] that we often fail to do what we ought to do (in this sense) is depressingly easy to make sense of. By contrast, transparent and reflective incoherence is hard to make sense of. When we are attributing transparently and persistently incoherent mental states to an agent,

---

[109] Cf., e.g., Davidson (2004: chs. 11-12); Dennett (1971).

[110] This may sound rather Kantian. The difference is that the Kantian thinks that the preconditions of agency are much more demanding than they may first appear, going beyond the relatively austere kinds of requirements I argue for in this dissertation. I'll have a bit more to say about this Kantian view in section 2.3.

[111] *Pace* the Kantian view mentioned just mentioned in the previous footnote.

we have reason to revisit our attribution and ask whether it is really the right way to make sense of her mental states.

Let me illustrate with an example: the requirement of instrumental rationality, which requires you to intend the believed necessary means to your ends. Here is a violation of the instrumental requirement which seems perfectly possible. You know that your friend's partner is cheating on her, and that she will discover this soon. You believe that it would be better if she heard it from you, both for her and for you (since she will also find out that you knew). So you intend to be the one who tells her about the infidelity. You also know (and hence, believe) that today is the last day on which you have the opportunity to tell her, and that the only way to do so is to call her. But, subconsciously motivated by the awkwardness of calling her, you put this out of your mind. You never form the intention to call her today, and you miss your chance.

What is harder to imagine is this: you are transparently aware of your intention, and your means-ends belief, and you say: "I intend to be the one to tell my friend about the infidelity. And the only way to do that is to call her today. But I have no intention at all to call her today." The most natural way to hear this speech is as a joke. Of course, it is possible that you do sincerely utter these words, but if so, you are confused about what it is to intend, and hence misclassify your own mental states. If, transparently to yourself, you really have no intention to call your friend today, even though you consciously recognize that this is the only way to be the one to tell her, then you just don't, in fact, count as intending to tell your friend. At best, you count as thinking that it would be good if you did so, or desiring that you do so, or something similar. This shows how the coherence requirement of instrumental rationality

(as opposed to the substantive reasons for action that one has) puts a constraint on our interpretation of you.[112]

Relatedly, coherence, rather than reasons-responsiveness, is the notion of rationality that has currency within social sciences such as economics, in their explanatory and predictive capacities. To try to explain human behavior on the assumption that people even approximately do or believe what they have most reason to do or believe would be utter madness. We fail to do that every day when we fail to be adequately kind, to give enough money to charity, to plan adequately for the future – and, yes, when we respond to our evidence imperfectly.

At the same time, we need notion of rationality that humans do at least tend to approximate, on the basis of which we can predict and explain their behavior. That property is coherence. This is why utility functions in economics are constructed out of bare preferences, rather than attempting to track the well-being that generates reasons for action in the philosopher's sense.[113] Likewise, the axioms of decision theory, which state constraints on rational combinations of preferences, are well-understood as coherence requirements. Philosophers have, I think, neglected the explanatory role that attributions of rationality play in the social sciences. This has caused them to overlook the interest and a notion of rationality as coherence, and to construct theories of rationality (in terms of reasons-responsiveness) that

---

[112] A parallel contrast holds about the attribution of beliefs. Any attempt to reinterpret belief so that only mental states that are genuinely responsive to the evidence count as beliefs will be implausibly restrictive and *ad hoc*. But to interpret beliefs so that they must conform to *coherence requirements* (under conditions of transparency) is much less restrictive, and much less *ad hoc*. Some read Davidson, a classic interpretationist, as holding that we must assume that others are responsive to substantive reasons to attribute mental states to them (cf. Shah & Velleman 2005: 516); but Davidson himself, at least in his later work, holds (correctly, in my view) that it is only rationality as *coherence* that must be assumed. See Davidson (2004: ch. 11), and also Blackburn (1998: ch. 3).

[113] Some economists conflate these, and in so doing needlessly commit themselves to a preference theory of well-being or to a crude internalism about reasons. But philosophically sophisticated economists recognize the distinction. For further discussion see Sen (1973); Broome (1991); Anderson (2001).

undercut its explanatory function. One virtue of understanding rationality as coherence is that it offers some prospect of unifying the normative and explanatory functions of rationality-attributions.

It is absolutely crucial to see that I am not claiming here that it is impossible to violate coherence requirements. It is perfectly possible to violate coherence requirements. My claim is that such violations are made possible by at least some degree of *non-transparency* of the agent's mental states to herself. There are many ways for your violation of coherence requirements not to be transparent to you. You can be in denial about your own attitudes – so that you sincerely deny that you have some intention or belief that you do in fact have. This may be due to your suppressing your recognition that you have the conflicting attitudes in question, or due to not realizing at any level whatsoever that you have these attitudes. Or, you may succeed in simply ignoring the conflict so that you never feel pressure to revise your attitudes. Or, even when you feel such pressure, you may be able to resist it simply by putting the matter out of your mind. In all of these cases, you are irrational, since you are irrational when you have mental states that violate the requirements of rationality, not just when you *think* that you have mental states that violate these requirements. So there is plenty of room to violate the requirements of rationality. What is hard is a stable, clear-eyed recognition that one has the attitudes that do so.

So deviations from coherence are still possible. Indeed, much recent social science – especially in psychology and behavioral economics – has been devoted to highlighting such deviations. Nevertheless, they are *deviations*, and they call out for explanation; indeed, that is precisely why uncovering the heuristics and biases which produce such deviations is such an intellectually interesting, and at times surprising, project. There's no such thing as the "failing-to-do-what-one-has-most-reason-to-do" bias, because one would never *expect* humans to do

what they have most reason to do. Deviations from coherence, by contrast, do call out for such explanation.

I hope this helps to explain why I take coherence to be of philosophical interest, and how a notion of rationality as coherence has an important role to play in philosophical theorizing. One might still worry that coherence, as distinguished from reasons-responsiveness, is not of distinctively *epistemological* interest. I tried to give one illustration of how coherence is epistemologically interesting in chapter 1, by showing how so-called "level-bridging principles", relating one's first-order beliefs and one's higher-order beliefs about what one's evidence supports, are best understood as coherence requirements, and not as claims about reasons. Chapters 4, 5 and the conclusion will give further examples of places where coherence requirements have an important role to play in epistemology.

2.2 The normativity of coherence requirements?

Having now said something about the philosophical importance of coherence requirements (for theorists), we are left with the question about their normative importance (for agents). Should an agent care about being coherent? There is an increasingly sophisticated debate on this question – sometimes called the debate on the "normativity of rationality" – in the literature on practical rationality.[114] Though I am unable to enter into it in detail here, I will make some remarks that bear on it.

It is not always clear what it means to ask whether coherence requirements are normative. On one relatively broad reading of 'normative', one that contrasts with 'descriptive', it is just obvious that coherence requirements are normative. They make claims

---

[114] For just a sampling of a growing literature, see Kolodny (2005); Raz (2011: ch. 8); Broome (2013: ch. 11); Southwood (2008).

about what is required of agents, and the notion of something's being required is a normative one, in this sense. So far, that is a fairly weak claim. In this sense, the 'ought' of 19th century British etiquette, or of Mafia morality, is also normative. One can felicitously speak about the requirements of British etiquette, or of Mafia morality; these requirements are normative in the sense that contrasts with 'descriptive'.[115]

There is a somewhat stronger sense, however, in which coherence requirements are normative. As I said in the previous section, to call someone irrational is clearly to express a kind of criticism: it is not, in this sense, a normatively neutral claim.[116] Indeed, as I also said above, this criticism can often get a motivational grip on the agent that is being criticized better than an accusation of imperfect reasons-responsiveness can, since it suggests some kind of failure by the agent's own lights. This is not true of an accusation that someone has violated the requirements of 19th century British etiquette. Such an accusation need not express criticism, and can easily be utterly inert in getting a motivational grip on the agent to change her behavior.

That said, if the reason that an accusation of incoherence gets a grip on an agent is that it suggests a failure by the lights of her own *substantive* normative standards, then this suggests that the reason the accusation of incoherence gets a grip on an agent is not because the agent cares about coherence *per se*, but rather because the agent cares about conformity with her own substantive standards, and the pursuit of her own substantive ends. This point

---

[115] One might say that these requirements are not, in fact, *genuine* requirements on agents – that is, that agents are not in any good sense genuinely required to comply with them – whereas coherence requirements are genuine requirements. I am open to that suggestion, and if it is right, I think it relates to the point I make in the next paragraph.

[116] One might distinguish the evaluative and the normative and argue that talk about rationality, and criticism of irrationality, can be construed as *evaluative* rather than normative. This is the approach that Easwaran & Fitelson (2015: 63) take. I want to at least leave open a more robust normativity for rationality-talk, as we'll see below.

needs to be managed carefully. As I argued in chapter 1, it actually is possible to be incoherent without failing to respond to one's own substantive reasons. But as I also argued there, in these cases one is not in a position to know that one is successfully responding to one's substantive reasons; on the contrary, it will seem, from one's point of view, that one is failing to respond to them. So even in such cases, the motivational grip of the accusation of coherence can be expected to persist.

This picture suggests that, at least typically, coherence is not itself the *end* of deliberation: it is not something that the agent aims at in and of itself.[117] As I argued in the instance of (ILC) in section 1.6, there is something odd about the idea that considerations of coherence weigh directly against the reasons to take that attitude in deliberation: that is, about thinking (for example) "well, on one hand, believing p would make me incoherent, but on the other hand it's what I have most reason to believe…"[118] But coherence nevertheless has a broadly normative role to play in that one is guided and constrained in one's deliberations by a disposition to conform with coherence requirements. From one's own, first-personal

---

[117] In this respect I agree with Kolodny (2005, 2007).

[118] There is a complication here. It is easy to imagine cases where one can experience, from a first-personal perspective, conflicts between coherence requirements and a certain particular kind of (purported) reason. Just take a coherence requirement, and then imagine that some demon or evil genius offers you a huge reward for violating it, or threatens some great harm if you don't. It seems perverse to think that keeping one's mental states perfectly coherent is more important than saving one's own life, or that of one's family, or whatever else one may imagine the payoff of the violation to be. So, one might think, it's natural to say – if we set the case up right – that one's reasons on balance support having attitudes that violate the coherence requirement in such a case. However, if there are indeed reasons to have the attitudes that violate the coherence requirement in this case, then they are so-called "state-given" reasons – roughly speaking, reasons that arise from the goodness of being in the state of having the attitudes in question, and not from considerations related to the object of the attitude and the fittingness of the attitude to this object. These latter are "object-given" reasons. It is a matter of debate whether state-given reasons are genuine reasons for attitudes (or whether they are just reasons to do things that get one to have those attitudes). (For relevant discussion, see amongst others Shah (2006), Parfit (2011: Appendix A), Piller (2006), Reisner (2009), and Schroeder (2012).) But if there are, they are, notoriously, reasons that do not effectively enter into *deliberation* about what attitude to have in the normal way. So although these cases might be experienced as a conflict of coherence and reasons, they are not experienced as a conflict between coherence and reasons *in one's deliberations* about the same unified question of what to believe. See also Hieronymi (2005).

perspective, it seems like a failure to conform with coherence requirements involves one's going wrong in some more substantive way. So the disposition to conform to coherence requirements will seem like one that is worth having, and indeed worth manifesting on individual occasions.

This is close to what Niko Kolodny (2005, 2007, 2008a, 2008b), a prominent "myth theorist" (that is, a denier of the claim that rationality understood as coherence is normative), has called the "transparency account". However, Kolodny goes slightly further. His view is that there is no *reason* to be coherent *as such*, but that the states that one has most substantive reason to be in will always have the property of conforming to the (purported) coherence requirements, because responding to one's reasons guarantees coherence.[119] If the argument of chapter 1 is correct, however, this is actually not right. It looks tempting for some coherence requirements, notably the noncontradiction requirement: it is hard to see how one's reasons could actually support both believing p and believing not-p. But it does not generalize to coherence requirements more broadly. Specifically, it fails for violations of (ILC). In cases of iterative failure, responding to one's epistemic reasons will make one incoherent.[120] So I reject Kolodny's transparency account as stated.

---

[119] A complicating factor in reading Kolodny is that in his later papers, he says that he is arguing against requirements of "formal coherence as such," whereas in the earlier (2005) paper, he clearly thinks that there *are* rational requirements, in a substantive sense where he takes a real stance on what they are (see e.g. 2005: 542), but that there is no reason as such to comply with them. This may be a change in his view. But it could alternatively be that the "requirements of formal coherence as such" that he is referring to later, and argues against (he focuses on noncontradiction, closure, and means-ends consistency) are not supposed to include the requirements that he does think are requirements in the earlier paper (which, by contrast, are requirements to believe and intend what one believes one ought to believe and intend).

[120] A similar conflict can be constructed for the (practical) enkratic requirement. If one can have sufficient reason not to Φ while also having sufficient reason to believe that one ought to Φ, then one can violate the enkratic requirement without going against one's (other) reasons. Indeed, I suspect that similar conflicts can be generated for any requirement that takes wide-scope over a combination of states that include a belief.

Since I reject the transparency account, I also want to leave it open that there is at least in some good sense genuine reason to comply with coherence requirements, even in the case where they come into conflict with one's substantive (viz., epistemic, moral, prudential, etc) reasons.[121] Whether one wishes to endorse this claim will depend on one's conception of reasons. In particular, if one thinks that coherence can only be a source of reasons for an agent if coherence itself is what the agent herself aims at in deliberation, then this suggests that one should reject the claim that coherence can be a source of reasons. But I do not want to take a stance on this. Note also that the idea that there are reasons to comply with coherence requirements is not at all incompatible with the idea that I have been stressing continually that coherence requirements are not themselves about responding to the totality of one's reasons. Coherence might be one source of reasons, compatibly with there being many other reasons that it is not incoherent to fail to respond to. So I don't say that my view rules out the possibility that there are reasons to be coherent.

Admittedly, one can certainly sensibly ask what the distinctive, exceptionless, non-instrumental *value* of being coherent is, as both Kolodny and Joseph Raz, the other most prominent myth theorist, do. They concede that a general disposition to be coherent is likely to be valuable to an agent over the long-run, but make the familiar point that this does not entail that individual manifestations of that disposition have any value.[122] For this to support a myth theory, there has to be some kind of connection between reasons and value. For example, we might say that for there to be a reason to Φ, Φ-ing must realize (or be expected

---

[121] On this view, the reasons to be coherent are a *sui generis* category of reasons, distinct from these other categories. This appears to be the view that Broome (2013: ch. 11) accepts, though he admits that he has no knock-down argument for it.

[122] See Kolodny (2008a: 459-60); Raz (2011: ch. 3). There may be room for querying this assumption, in my view, but I have no space to do that here.

to realize) some value. This is a suppressed premise in the argument, which is needed to derive the conclusion that since being coherent does not realize any value, there is no reason to be coherent.

But as Raz himself (2011: 41-5) argues elsewhere, the connection between reasons and value is arguably severed in the case of *epistemic* reasons. In just the same way that we can ask, "what's the value of being coherent *as such*?," we can ask "what's the value of believing the truth *as such*?"[123] Both dispositions to be coherent and dispositions to believe the truth are valuable over the long-run, but in both cases the question of the value of individual manifestations of these dispositions arises. And I do not particularly see why the question is supposed to be more urgent in the former case than the latter.[124] If we concede that there can be epistemic reasons that are not tied to value, then that seems to leave it open that there could be reasons of rationality that are not tied to value. The argument from the lack of a distinctive value of being rational as such to the absence of reasons of rationality would then fail.

Moreover, I am not fully convinced, even if it weren't right to talk of distinctive *reasons* of rationality, that coherence requirements might not be construed as genuinely normative in a good sense. Perhaps coherence requirements are just a fundamental kind of *normative* requirement, the normativity of which is not to be understood in terms of reasons,[125] with two fundamental (and perhaps incommensurable) normative categories here.[126] Or perhaps

---

[123] Indeed, some philosophers have seriously questioned the normativity even of morality (cf., e.g., Copp 2007: ch. 8).

[124] Of course, one can simply posit that believing the truth has intrinsic value and leave it at that. But two can play at that game: one can simply posit that being coherent has intrinsic value and leave it at that.

[125] Nic Southwood (ms.) is working out a view like this.

[126] Though perhaps some specialized notion of a 'reason' could itself be defined in terms of responding to coherence requirements, as Ralph Wedgwood suggested to me.

coherence requirements are construed as having an impact on an agent's normative position in some more complex sense. For example, on one proposal, in cases of conflict between coherence requirements and substantive reasons, the fact that one cannot coherently respond to one's reasons in some way cancels the ordinary normative force that those reasons carry, or at least makes one's failure to respond to them blameless.[127] These are all interesting possibilities. But I cannot do justice to all of these possible views here. I simply conclude that we have no decisive argument that coherence requirements are not normative in a robust sense. Moreover, even if coherence requirements are not normative in this robust sense, they are clearly normative in various weaker senses – and, as I argued in the previous section, of distinctive philosophical interest that is unconnected with their normativity. This is enough for my project.

2.3 The centrality of coherence to the ordinary concept of rationality

To reiterate, however one resolves the questions surveyed in the last section about the normativity of coherence requirements, the result argued for in chapter 1 and in section 2.1 stands: coherence requirements are distinct from substantive reasons and are not to be understood in terms of responding to the totality of one's substantive reasons. I already offered some initial arguments, in sections 1.6 and 1.7, that in the light of possible conflicts between reasons-responsiveness and satisfying coherence requirements, we should use the term 'rationality' to cover the latter only. I want to now bolster these arguments by bringing out how coherence – in the sense expounded through the dissertation so far and especially in section 2.1 above – is central to our ordinary notion of rationality.

---

[127] Thanks to Geoff Sayre-McCord for these suggestions.

My approach will be to contend that incoherence is a deeper and more fundamental kind of irrationality than imperfect reasons-responsiveness. My claim is not that the ordinary term 'rational' is never used to talk about failures to respond to reasons. Rather, it is that it is *more* central to our ordinary notion of irrationality that incoherence should always count as irrational than it is that failures to respond to reasons should always count as irrational. In fact, I will try to elicit the intuition that in a core sense of 'rational', failing to perfectly respond to one's reasons is *not* irrational. I will argue that this is true in both the practical and doxastic cases, but I will admit that the intuition is much stronger and easier to generate in the practical case than in the doxastic case. I will devote much of the rest of the chapter to the interrogation of this asymmetry.

Let us begin then, with the moral case, and with the way in which the following two claims at least often carry different connotations:

(1) You really ought not to have done that!

(2) Doing that was irrational!

Return to the example of eating meat, which I brought up in section 2.1. Suppose that I think that eating meat is wrong, and that this moral fact gives others decisive reason not to eat meat. Of someone who has just eaten a large veal steak, I suggest, it would be very natural for me to utter (1), but much less natural for me to utter (2).[128] Even as I judge that (say) eating meat is morally wrong, and that some agent has decisive moral reasons not to do so, it will still seem wrong to call her *irrational* for eating meat. That is not just out of politeness: I am perfectly

---

[128] See also Ridge (2014: ch. 8, esp. 231-33).

willing to say that she has failed to do as she ought. Rather, the charge of irrationality just somehow seems off-base. Specifically, it seems particularly off-base if the agent herself does not share my moral views, and believes that eating meat is perfectly permissible.

Why is this? As I imagine the case, it is not that I think that even though meat-eating is wrong, the agent has other, non-moral reasons – such as the self-interested reasons generated by her own enjoyment – that trump these moral reasons.[129] I think the moral reasons are decisive. So, this is not what explains my reluctance to call the agent irrational. I do not think this is a quirk on my part: often, when people judge that others are doing moral wrongs, they take those moral wrongs to be serious and weighty, and not to be normatively trumped by considerations of self-interest. Still, they show reluctance to describe immoral behavior as irrational.

Instead, I think that what explains the "off-base" feeling of the charge of irrationality is that, as described, the agent is not incoherent in any way. Her acts are concordant with her own moral standards, and as such she may still be rational. This isn't necessarily to embrace a subjectivism about morality itself, or its reason-giving capacity; it is to embrace coherentism about *rationality*.

By contrast, consider another case.[130] Suppose that you don't even *like* meat. Rather, you much prefer eating vegetarian dishes. However, you have been unable to make up your mind between two delicious vegetarian dishes, both of which you love – quinoa with quince jam, and kale with kumquat. Paralysed by indecision between the two, you find yourself ordering a third option: veal with a foie gras foam, even though you would prefer either of the

---

[129] Nor need it be that I am a "reasons internalist", and think that agents only have reasons to comply with moral demands if they desire to comply with those demands. For discussion of reasons internalism, see 2.4 below and 0.3 above.

[130] Kamm (2007: 485-6) attributes a similar case to Sidney Morgenbesser.

vegetarian options to this veal dish. Now it's extremely intuitive to call your choice irrational. This is because there is a kind of internal incoherence in your total set of preferences and choices. In this case, it's very natural to say that your decision to eat meat is irrational, because it involves you in incoherence. The distinctive criticism expressed by calling someone irrational here seems to be exactly the distinctive criticism of calling someone incoherent that I identified in section 2.1.

The point is more general, I think. Quite generally, when someone acts wrongly, but does so in a way that he himself does not (at any level) recognize as wrong, it is very natural to tell him that he ought to have done otherwise, or that he was unreasonable, but less natural to accuse him of *irrationality*. Consider the range of different societies in which individuals have grown up over human history. Given the conflicts in moral beliefs across this range of societies, it is hard to avoid the claim that if there are normative facts, then some of the societies have been wrong about some of them. But it would be bizarre, I think, to call all those who have lived by their own social moral codes *irrational*. I do not think we would be ordinarily inclined to make this charge.

This pattern persists even when we imagine someone committing an uncontroversially serious moral wrong. For example, suppose a politician murders an aide in order to cover up his own involvement in a political scandal so as to save his career. This action, I take it, we would all agree is seriously wrong, and I think most would agree also that the politician's self-interested reasons to kill the aide do not outweigh his moral reasons not to do so. But far fewer of us, I suggest, would necessarily be inclined to describe the politician as irrational for committing this act. Of course, we could develop the case in various ways such that we would be more inclined to describe him as irrational – for example, if the politician himself judges that he ought, all-things-considered (not just morally), to refrain from killing the aide, or if

doing so ultimately frustrates his own ends. But again, such stipulations make the politician incoherent. If it is stipulated that the case is *not* one in which the politician is incoherent in one of these ways, then there are many things we might say naturally against him – but accusing him of irrationality is not high on the list.

Many moral philosophers, I think, would concur with these intuitive verdicts: failures to respond appropriately to one's moral reasons do not necessarily constitute irrationality. However, some may protest that this is too quick. Isn't this one of the great debates in ethics – that over the truth of so-called "moral rationalism"? Numerous thinkers – notably those in the Kantian tradition – have claimed that, after all, moral failings *are* rational failings.

In response to this, let me make two points. First, I want to reiterate the point I made in the introduction (specifically, section 0.3), namely that distinguishing rationality and reasons is itself helpful in this debate. Historically, the term 'moral rationalism' has been used to refer to two quite different claims. The first claim is that morality is (contra so-called "reasons internalism") a categorical source of practical *reasons* for agents – that is to say, agents have practical reasons to comply with the true morality, irrespective of whether they accept this morality, or desire to comply with it. The second claim is that agents are *rationally required* to comply with the demands of morality. It is only if one equates rationality with responding to reasons that these two claims appear to be the same. Much of what many so-called "moral rationalists" wanted can be granted by agreeing that morality is a categorical source of *reasons* for agents. That is not something that we need to deny in claiming that failures to respond correctly to one's moral reasons are not necessarily irrational.[131]

---

[131] Let me also stress a point which should be obvious, namely that to say that it is not necessarily irrational to be immoral is certainly not to say that it is (even in some cases) irrational to be moral. If one starts with the idea that rationality is about responding to reasons, one is quite naturally led to the thought that, since the balance of reasons usually (perhaps always) supports one unique action, there is no gap between what is rationally

That said – and this brings us to the second point – there will be those who hold, even after the distinction between reasons and rationality has been drawn, that failures to respond to one's moral reasons really do ultimately always amount to rational failings. However, note that in general those who argue this do not take it as axiomatic or analytic – rather, they take it as a claim to be argued for. More particularly, recent "Kantian constructivists" like Christine Korsgaard want to show us that ultimately, failure to comply with the moral law amounts to a kind of internal incoherence between one's (intended) actions and the universal principles that one implicitly commits oneself to through volitional acts.[132] This shows that even these moral philosophers have a concept of rationality that is closely tied to coherence. Their (ambitious) task is to try to show that failures to respond to moral reasons ultimately involve the agent in incoherence. I do not think such attempts succeed, but I agree with them that this is what would need to be shown in order to demonstrate that moral failings are necessarily rational failings.

So, all in all, I think that the moral case exhibits the centrality of coherence, and the relative marginality of reasons-responsiveness, to our ordinary concept of rationality. Of

---

permitted and what is rationally required. So, the thought would go, if there are situations where it's rationally permitted to deviate from morality, then those must be cases where it is rationally required to deviate from morality, and consequently must also be cases where it is rationally impermissible to be moral. Moreover, on this view, one will have to say that the only way that it could be rational to be immoral would be if one's non-moral reasons are outweighed by other reasons. (The most vulgar view in the neighborhood is the view that egoistic reasons always outweigh moral reasons, which combined with the view that rationality involves responding to reasons yields the dreaded "rational egoism." I think this is one of the worst, and most pernicious, views of rationality going. In accepting that one may be immoral without being irrational, I do not in any way mean to express sympathy for a view like rational egoism, which is a particular kind of reasons-responsiveness view, not a coherentist view.) But if rationality is not about responding to reasons, then it is much more natural to think that rationality permits multiple courses of action. So it may well be that cases where one can deviate from morality without irrationality are not cases where complying with morality would be *irrational*. Moreover, in saying that there are cases where one can be immoral without being irrational, the coherentist makes no commitment to the claim that in such cases non-moral reasons outweigh one's moral reasons (indeed, perhaps they *never* do).

[132] See, e.g., Korsgaard (2009). Other examples of attempts to execute such a project (in very different ways) include Sterba (2013) and Lord (ms.).

course, I am not claiming that the English word 'irrational' is never used to talk about agents failing to comply with their moral reasons. No doubt the term 'rational' is used in a variety of different ways.[133] In arriving at an adequately unified theory, some inconsistency with ordinary usage is inevitable, since ordinary usage is not perfectly unified. As I said in section 1.7, we have reason to prefer a terminology that avoids running distinct phenomena together or leaving some of them out of our conceptual scheme altogether, and using 'rational' to refer only to the satisfaction of coherence requirements achieves this well, in a similar way that the philosophical use of 'logical' does so.

Recall that it won't do to use 'rational' to cover both coherence and evidence-responsiveness, since the two come into conflict. Moreover, the intuitive data surveyed in this section suggest that often, speakers are actively *disinclined* to use describe failures to be reasons-responsive as 'irrational'. It is not that people consistently use 'rational' to refer to a disjunctive property of coherence and reasons-responsiveness – thus sometimes applying it to incoherent agents, and sometimes to reasons-unresponsive agents. Rather, they often use it in a way that actively excludes reasons-responsiveness.[134]

---

[133] Unlike, say, 'knowledge', 'rationality' does not seems to be a lexical universal. There are languages such as Latin that have several conceptually distinct words (*sanus*, *sapiens* and *rationalis*) that are each sometimes translated as 'rational'. The English 'rational' comes from the Latin *rationalis* and the fact that we have the distinct English words 'rational' and 'reasonable' probably reflects that fact that Latin, from which English got the term, slices the conceptual space more finely in this area than many other languages do. Other Latin-influenced languages such as French and Italian also distinguish 'rational' and 'reasonable'. By contrast, German has *vernünftig*, which is indeterminate between 'rational' and 'reasonable', plus *rational* as a recent English import. Turkish is very similar, with *akılcı* for either 'rational' or 'reasonable', plus the English import *rasyonel*. And Irish and Japanese each have no original word for 'rational', just words for 'reasonable' (*réasúnach* and 合理的 [*gouriteki*] respectively; Japanese now has the English import ラショナル [*rashonaru*]). There are also languages where the closest word for 'rational' can also be translated as something else that has an entirely different meaning from either 'rational' or 'reasonable' in English. For example, in Mandarin Chinese, 有道理 (*yǒudàoli*) can mean 'fair' or 'virtuous' and is etymologically related to 'the natural (or universal) way)'. For providing me with many of these examples, I'm grateful to Jonathan Leader Maynard, Thomas Mallon, Deirdre Milner, and Yuan Yang.

[134] As Keith DeRose pointed out to me, this leaves it open whether there might be some other account of rationality or of uses of 'rational' that involves both conformity to coherence requirements and at least *some*

So far in this section, we have been focusing on the case of the relationship (or lack of) between moral reasons and rationality. This has not been an accident: I think it is the case which most strongly elicits the centrality of coherence to our ordinary concept of rationality. That said, in this dissertation, I am focusing primarily on rational belief. So we must face up to the case of epistemic reasons. I have been focusing on the moral case to make the strongest case for the point that ordinary usage does bear the imprint of a difference between the notion of what one ought or has most reason to do on one hand, and what it is rational to do on the other. I cannot deny that this difference is less marked in the epistemic case. *Prima facie*, I think we have reason to want continuity between our theory (and vocabulary) of practical rationality and our theory (and vocabulary) of doxastic rationality. So I take the examination of the moral case to be relevant in one way already. But I need to say much more.

In the remainder of this section, I will try to make the case that our ordinary usage does bear *some* imprint of the difference between reasons and rationality even in the case of epistemic reasons. However, I will still conclude that there is a genuine asymmetry in linguistic

---

responsiveness to reasons. For example, it might be that we are inclined to describe the most egregious failures to respond to reasons as irrational, but not to describe just any old failure to respond to reasons as irrational. Or perhaps we have some sense of which reasons are really "mandatory" for us to rationally respond to, and which are not. The challenge, of course, is to show that there is a good way of drawing the line between the failures to respond to reasons which do count as irrational and those which do not, that leaves us with a notion of rationality that is of genuine theoretical interest, and which keeps distinct things apart to an adequate degree. It is not clear why one, as a philosopher, should be particularly interested in a notion of rationality as coherence-plus-no-egregious-failures-to-respond-to-reasons, rather than simply theorizing coherence requirements on one hand, and the whole gamut of reasons on the other.

Setiya (2004) offers an intriguing view which offers some promise of meeting this challenge. Setiya agrees that failures to respond to reasons are not necessarily irrational, but does not theorize rationality as pure coherence. On his view, roughly speaking, one is irrational in holding some attitude when one could have been legitimately *expected* not to hold the attitude. This seems in the ballpark of delivering what DeRose suggested to me: plausibly, failures of coherence will always be violations of what can be legitimately expected of me, but only the most egregious failures to respond to reasons will violate such expectations. For example, when I fail to respond to my reasons because I have deeply-held but false normative beliefs which I am acting on in good conscience, I am not violating legitimate expectations (at least in Setiya's own view). Yet still, the question is whether the notion of a legitimate expectation is really something that we can get a prior fix on independently of rationality, such that the former notion can truly illuminate the latter. I am skeptical about this, and still find the coherentist approach cleaner. Still, I think that Setiya's view is a serious contender which deserves more consideration than it has received in the literature to date.

behavior and conceptual intuitions here. In section 2.4, I will be examining whether this asymmetry can be vindicated. Ultimately, I will offer an error theory for it, in section 2.5.

Here is a case in which it seems that our talk about what it's rational to believe on one hand, and that about what the evidence supports on the other, can come apart quite naturally. Suppose that someone grows up in a cult that teaches her that if she holds anyone's hand, she will get a deadly disease and die. I think many people would be inclined to say that the testimony of the members of the cult is evidentially pretty worthless: though the person (understandably) takes it to be good evidence, it really isn't – not, at least, if the cult members are sufficiently unreliable with respect to these matters generally. So the testimony isn't sufficient to make it the case that her evidence supports believing that if she holds someone's hand, she will get a deadly disease and die. (Maybe it only supports suspending judgment rather than disbelieving this proposition, but that's fine.) Still, I don't think we'd be as inclined to describe her as irrational for having the belief she does given her upbringing in the cult.

To reinforce this, consider how she might herself react upon discovering that her views are mistaken. It'd be very natural to say "it turns out that, though I thought I had good evidence for my belief, really I didn't: that testimony I was given was worthless; every word they say about this kind of thing is false." (As opposed to, say, "although I did have excellent evidence for my belief, I have now received a defeater for this evidence, and no longer have excellent evidence for it.") But it'd be less natural, I think, to say, "it turns out that, though I thought I was rational, really I wasn't." If the ex-cult member said this, I think we'd think she was being too hard on herself. She didn't know how crummy her evidence was. Here, ordinary language does seem to recognize at least some distinction between rationality and responding to evidence.

What about in the philosophical discipline of epistemology? Here, I think, there has been some limited recognition even in epistemology that incoherence is at least a deeper and more fundamental kind of failure of rationality than insensitivity to evidence. Recall from section 1.3 Elga's description of his friend Daria. He describes her belief in astrology against the actual evidence as "less than perfectly reasonable." But he describes her so believing in the face of her own judgment about her evidence as "an insult to rationality". I do think that there is a solid intuition here that the latter case involves irrationality in a deeper and more fundamental way.

That said, there is undeniably a very natural tendency to describe beliefs founded on what is obviously just very poor evidence as 'irrational', irrespective of their satisfaction of coherence requirements. So, for example, if I judge that some agent's evidence conclusively supports disbelieving the proposition that there are fairies living at the bottom of her garden, then I might quite naturally describe her as irrational for believing this proposition, even if she herself believes that her evidence supports it. Indeed, ordinary parlance lacks natural terminology for distinguishing an epistemic failing from a rational failing – since, unlike 'moral failing', 'epistemic failing' is not a folk locution. When it comes to belief, 'irrational' can seem like the only ordinary term of criticism available.

We find, then, ourselves with a striking asymmetry. Whereas it is natural to think that failures to respond to moral reasons do not (in and of themselves) amount to irrationality, it is also standard to think that at least severe failures to respond to epistemic reasons do amount to irrationality. Again in line with the "disambiguating response" of section 0.2, one might propose here that this asymmetry is just a matter of divergent uses of the term 'irrationality' to refer to incoherence or to reasons-unresponsiveness, and that by pure chance it has apparently evolved to refer largely to the former in the moral case, and largely to the latter in

126

the epistemic case. But – and I am now explicitly tackling the issue that I bracketed in section 0.2 – in many ways this seems unsatisfying. If the asymmetry were only found in academic philosophy, we could perhaps countenance the possibility that it is just a chance sociological fact that epistemologists have used 'rationality' in one way, and ethicists have used it in another.[135] But the asymmetry is also found in ordinary talk and judgments, by people who are neither epistemologists nor ethicists. This suggests at least *prima facie* that there may be some tacit set of philosophical presumptions that explains the asymmetry. It is thus worth trying to identify what these presumptions might be, and whether they can be vindicated, or explained away.

2.4 Attempts to vindicate the asymmetry

I can think of four ways that philosophers might try to *vindicate* the asymmetry that I have been calling attention to.

(a) *Differential access to the relevant reasons*

The first way is by claiming that whereas our moral reasons may in some important sense often not be (epistemically) available to us, our epistemic reasons always are (epistemically) available to us. So, if rationality is about responding to the reasons that are *available* to us, our asymmetry might be explained.

---

[135] Much as, for example, economists typically use 'utility' to refer to preference-satisfaction, whereas moral philosophers typically use it to refer to well-being (which might or might not turn out to be best understood in terms of preference-satisfaction, depending on one's *substantive* theory of well-being). See Broome (1991).

The crudest version of this line of thought suggests that one's evidence must be something about which one, in principle, cannot be misled in any way.[136] This rests on an extreme version of "internalism" about evidence; I illustrated one way in which this view goes wrong in section 1.4.

Still, there may be a lingering suspicion that in some way epistemic reasons have to meet some sort of accessibility constraint that moral reasons do not. I think this thought rests on an equivocation that can be made clear by focusing on an important distinction between different sources of epistemic ignorance about your reasons. Suppose that some proposition R (or the fact corresponding to this proposition, if you prefer) is a reason to Φ. Your ignorance that R is a reason to Φ may be traceable to (at least) two quite different sources. On one hand, it could be due to descriptive ignorance of R's truth; on the other hand, it could be due to ignorance of the normative facts that make R a reason to Φ.[137] Let's take these cases one at a time to see if there is an asymmetry between moral and epistemic reasons with respect to either.

Begin with the case in which you are ignorant of R's truth. One might think there is an asymmetry here between moral (and more generally, practical) reasons on one hand, and epistemic reasons on the other. We are talking as if R can be a reason (for you) to Φ even if you are unaware that R is true. It may strike readers that this is plausible when R is a practical reason and Φ is an action, but not when R is an epistemic reason and Φ is a doxastic attitude.

---

[136] Foley (1993: 191-194) seems to presuppose such a view, and ends up with a very subjectivist view of evidence and evidential support. Strikingly, some philosophers have just used 'evidence' and 'what one takes to be one's evidence' quasi-interchangeably: see, e.g., Adler (2002: 8); Smith (2007: 280, 283); see section 4.5 for more on this.

[137] Here I'm using the schematic symbol 'Φ' to stand in not just for potential actions, but also for potential mental states or attitudes.

After all, an epistemic reason has to be an evidential reason, and R won't be part of your evidence unless you're aware of it.

However, this elides an important distinction. We can talk of evidence and reasons to believe that you are unaware of.[138] For example, we can say, "actually, there is very good evidence for the belief that climate change is man-made – Jeff just isn't aware of it." Or we can say, "actually, there are very good reasons to believe that Rashid will be resigning – Tamara just isn't aware of them." What is less natural is to ascribe *possession* of this evidence or reasons to the person unaware of them. So it's unnatural to say that Jeff *has* good evidence that climate change is man-made that he is unaware of, or that Tamara *has* good reasons to believe that Rashid will be resigning. So there can in a good sense *be* an epistemic reason R for you to Φ (where Φ-ing is the having of a doxastic attitude), but in a good sense you don't *have* this epistemic reason if you are unaware that R is true. I see no reason to treat the practical case differently. There can in a good sense *be* a practical reason R for you to Φ (where Φ-ing is an action), but in a good sense you don't *have* this practical reason if you are unaware that R is true.[139]

Turn now to the case where you are aware of R's truth, but not aware of the normative facts that make R a reason to Φ. Here, again, things are parallel. In the practical case, where R

<hr />

[138] See also Sylvan (ms.: 17).

[139] See Audi (2001: 53); Dancy (2000: 65-6); Skorupski (2010: 108); Smith (2007: 281); Woods (1972: 190). Schroeder (2008) criticizes this notion of 'having' a reason (see Lord (2010) for a reply). Schroeder, too, however, wants in his rival account to treat the doxastic and practical cases in parallel. So his criticism is not a threat to the broader claim I am making here.
  Keith DeRose pointed out to me that the ordinary locution 'has a reason' is more natural in cases where the subject isn't aware of the reason in the practical case than it is in the doxastic case. This seems to be right to me, but I am not sure what to make of it. When it comes to a stipulative distinction between what we as theorists will mean by 'there being a reason' on one hand and 'having a reason' on the other, I do not see how to draw the distinction in a way that treats the practical and doxastic cases differently.
  Going in the opposite direction, someone might conceivably claim that if you are unaware of R, it could not even be a reason *for* you to Φ. I think this goes too far. But again, whatever you think, there seems good reason to treat the doxastic and practical cases in parallel here.

is a putative practical reason and Φ-ing is an action, your ignorance of the normative facts that make R a reason to Φ does not preclude R from being a reason for you to Φ. But similarly, in the case where R is a putative epistemic reason and Φ-ing is the having of a doxastic attitude (and the relevant normative facts are evidential support relations), ignorance of these evidential support relations does not preclude there being an epistemic reason for you to Φ. This is actually essential to the judgment, from the previous section, that the person who believes in fairies fails to believe in accordance with his epistemic reasons. After all, he is ignorant of the fact that his evidence doesn't really support belief in fairies. But epistemologists would not want to say that he has responded properly to his epistemic reasons. In sum, then, once we separate out the different sources of ignorance, there does not appear to be a real difference between moral and epistemic reasons with respect to epistemic accessibility constraints.

Richard Feldman (1988) offers a somewhat different sort of argument that tries to carve out at least some kind of asymmetry of this kind. Feldman argues that while "objective" ethical justification is independent of an agent's cognitive position or perspective, even the most "objective" epistemic justification depends upon the agent's cognitive position or perspective. I find this contention unpersuasive. As Feldman acknowledges, it depends upon the assumption that there is no serviceable concept of (super-)objective epistemic justification co-extensive with truth. While I acknowledge that it might not be natural to call such a concept 'justification,' I do not think this choice of terminology on Feldman's part is innocent. We can use truth as a normative standard against which to evaluate beliefs using other terminology. For example, the notion of a *correct* belief is, as several philosophers have stressed,[140] a normative one, and a belief is correct iff it is true. Moreover, there is a corresponding sense,

---

[140] See, e.g., Wedgwood (2002b); Shah (2003); Gibbard (2005).

for example, I think, in which we talk about what you *ought* to believe against the standard of truth; that is one of the semantic values that 'ought to believe' can take, depending on context.[141]

Moreover, if we do stick with Feldman's terminology of 'justification', it seems to me that talk of 'ethical justification' is not happily independent of the agent's cognitive position or perspective either. Suppose that someone does the objectively right action *entirely by accident*; really, they were trying to do something quite dastardly. It's pretty unnatural to describe their act as *justified*. One might respond to this that the correct ethical theory won't allow the action to be objectively right unless it is done with a (justified) expectation that it is right – but that is just to concede that the agent's cognitive position or perspective matters.

I conclude that considerations of epistemic accessibility provide no reason to think that failures to respond to your epistemic reasons are failures of rationality, whereas failures to respond to your moral reasons are not.

(b) Simplicity and obviousness of norms

Now to a second attempt to vindicate the asymmetry. Some philosophers may have thought that the epistemic norms that determine your epistemic reasons are obvious, whereas moral (and other) norms that determine your practical reasons are not. The idea may have been that while differing normative perspectives can rationalize different actions, there are no differing normative perspectives to rationalize different beliefs, since the candidate epistemic norms are ones we are all committed to: believe the truth, and (since evidence is a marker of truth),

---

[141] See Gibbard (2005) for one account of the relationship between correct belief, truth, and 'ought'.

believe what the evidence supports. Perhaps, if one or both of these norms are "constitutive" of belief, we are all implicitly committed to them, whereas there is no similar constitutive norm of action.[142] That is enough to make failures to respond to our evidence irrational (perhaps because they are ultimately a form of incoherence), and thus to explain the difference between the doxastic and practical cases.

The problem with this argument is that these epistemic norms fail to come anywhere near exhausting normative enquiry about epistemic reasons. For, as I stressed in section 0.6 and subsequently, evidential support relations – which specify what evidence counts in favor of what doxastic attitudes – are themselves normative. Consequently, one can easily be failing to respond to one's epistemic reasons while nevertheless inhabiting a coherent (though mistaken) normative perspective – by having incorrect evidential standards. Inhabiting a mistaken normative perspective with respect to one's beliefs does not require one to knowingly reject the norm of believing what one's evidence supports (as would be incoherent), any more than inhabiting a mistaken normative perspective with respect to one's actions requires one to reject the norm of doing what one has most reason to do. Again, when we properly compare like for like across the doxastic and practical cases, the purported asymmetry disappears.

In response, someone might suggest that the idea is not supposed to be that the general norms of believing the truth and of believing what one's evidence supports are the *only* epistemic norms. Rather, it might be contended, the idea is that if we are all committed to

---

[142] Some philosophers have tried to defend the view that there is a constitutive norm of *action* – see, e.g., Korsgaard (2009). For a critique of this approach, see Enoch (2006).

believing the truth, we are *instrumentally* irrational in some particular case when we fail to respond to our evidence – since responding to evidence is a means to believing the truth.[143]

However, no credible account of instrumental rationality vindicates the following schema: S has goal G; Φ-ing is a means to goal A, so S is irrational if S fails to Φ. There are several problems with this schema, but the biggest is that S may not be irrational if S does not have any idea that Φ-ing is a means to goal A, or if S only knows that Φ-ing is a means to goal A under some description under which she is unable to identity what Φ-ing is. For example, if I aim to get to Amarillo by 7pm, but have no idea that the train on platform B is the (only?) way to get to Amarillo by 7pm, I am clearly not irrational for failing to get on the train on platform B. Moreover, even if I know that taking the train going to Amarillo is a way of fulfilling my goal of getting to Amarillo by 7pm, I may not be irrational for failing to take the train going to Amarillo, if I don't know which train is going to Amarillo.

Similarly, just because I aim to believe the truth does not mean that I am instrumentally irrational for failing to believe what my evidence supports, if I do not know what my evidence supports. Granted, I am irrational if I *know* that my evidence supports some doxastic attitude and fail to take it; that is because (assuming that knowledge entails belief) in such a case, I violate (ILC); that is, because I am incoherent; not simply in virtue of my failing to respond to my epistemic reasons.

(c) Moral reasons can be outweighed, but epistemic reasons can't be

---

[143] This is of a piece with the "instrumentalist" theory of epistemic rationality discussed briefly in section 0.7 above.

A third proposal for vindicating the asymmetry is that moral reasons can be outweighed by competing practical reasons, whereas epistemic reasons have nothing that can outweigh them. This might explain why it is always irrational to fail to believe what one's epistemic reasons support believing, but not always irrational to fail to do what one's moral reasons (alone) support doing, consistently with an account on which rationality is a matter of doing or believing what the *totality* of one's reasons support doing or believing.

This proposal can be quickly dispatched. As already argued in section 2.3, the intuition that one can be immoral without being irrational persists even in cases of very serious wrongdoing, where one's moral reasons clearly outweigh one's prudential reasons. Moreover, even if one had a view on which prudential reasons tend to outweigh moral reasons, it is enough to fell the present proposal that speakers who *do* think the moral considerations outweigh the prudential ones are disinclined to attribute irrationality.

(d) Reasons internalism

The final potential explanation takes a slightly different form to the first three. Whereas they tried to offer some reason to think that rationality requires you to respond to your epistemic reasons but not your moral reasons, the fourth explanation claims that morality does not give you reasons when you do not accept its demands.[144] One might appeal here to Bernard Williams' (1981) famous doctrine of "reasons internalism", according to which all of one's reasons must be grounded in one's own subjective motivational set. To illustrate, consider

---

[144] One could alternatively put the point by saying the moral "reasons" do not amount to genuine *practical* reasons when one does not accept their demands. I do not think that any difference here will matter in this context.

again the case of the meat-eater. The view would be that, although the meat-eater perhaps in some sense (morally) ought to refrain from eating, he has no reason to do so. For there is nothing in his subjective motivational set to get him from the considerations that putatively tell against his eating meat to an intention not to eat meat.

There are three problems with this line of argument. First, few moral philosophers accept Williams' strong form of reasons internalism, and consequently I do not think it is what drives their intuition that the meat-eater is not irrational. It is not unnatural to say that the suffering of animals gives you a reason not to eat meat, whether you recognize this reason or not. Your failure to be moved by it does not mean that it ceases to be a consideration against your eating meat. Once again (cf. the discussion of "moral rationalism" in section 2.3 above, and section 0.3 before that), the reasons-rationality distinction affords us the chance to embrace a moderate view here, on which morality can give us *reasons*, but it is not necessarily *irrational* to fail to comply with these reasons; this latter claim may give some reasons internalists enough of what they really wanted here.

The second and third problems both illustrate that even if reasons internalism were true, it would not adequately explain the difference in intuitions between the moral and epistemic cases. The second problem is this: reasons internalists typically concede that one can have a *de dicto* desire to comply with one's moral obligations, and that this is enough for the (actual) moral facts to give one reasons.[145] So let us just imagine our meat eater as having a desire to comply with his moral obligations. In that case, even by the reasons internalist account, he has a reason not to eat meat, if doing so is his moral obligation.[146] But I still think

---

[145] See, e.g., Railton (1986).

[146] The fact that he is *unaware* of this fact does not preclude him from having a reason on the internalist account. The internalist restriction is that the fact has to be capable of moving you *if* you are aware of it, not that you have to be aware of it.

we would not be inclined to call the meat eater irrational, if he thinks that he does nothing morally wrong by eating meat.

Third, consider again the example of the person who believes in fairies. This person is aware of the facts that (putatively) give him reason not to believe in fairies, and these facts do not move him to revise his belief that there are no fairies. So, by the reasons internalist view, those facts don't give him any reason not to believe in fairies. So the reasons internalist cannot count the fairy-believer as failing to respond to his reasons any more than she can count the meat-eater as failing to respond to her reasons. Of course, the fairy-believer isn't aware that the facts *are* reasons for him not to believe in fairies. But nor is the meat-eater aware that the facts (of suffering) *are* reasons for him not to eat meat. Again, once we adequately distinguish descriptive and normative sources of ignorance, there are still no grounds for a judgment of asymmetry here, even after introducing reasons internalism.

There is a final, general reason to be skeptical of all three strategies for explaining the difference in intuitive verdicts. The asymmetry we noted at the start was between failures to respond to one's moral reasons and failures to respond to one's epistemic reasons. Each of the strategies we have considered so far explains this difference in a way that, if it succeeded, would appear to ground some general difference between responding to epistemic reasons on one hand, and responding to practical reasons on the other. But such an explanation would only work if intuitions about failures to respond to other, non-moral practical reasons received the same kind of intuitive verdict as failures to respond to moral reasons.

In fact, that does not seem to be so. Consider, for example, Rawls' example of the person who desires and intends only to count the number of blades of grass in his garden.[147] Such a person plausibly fails in some way to respond to his non-moral practical reasons. And in ordinary discourse it's pretty common to describe someone like the grass-counter as irrational – even if we stipulate that he is ideally coherent.[148] The intuitive reaction to the case is more like that to the person who believes in fairies than it is like that to the person who eats (and enjoys) meat. So failures to respond to non-moral practical reasons seem to intuitively go along with epistemic reasons rather than moral reasons.

Given that, none of the putative explanations here seem even to attempt to explain the difference in intuitions that we actually find ourselves with. The grass-counter is failing to respond to reasons for action, rather than reasons for a doxastic attitude. Moreover, (if he is ideally coherent) he lacks a subjective motivational set that would move him not to count grass. So all of the explanations that we have considered of the purported difference between the cases would predict that the intuition about the grass-counter should go along with that of the meat-eater rather than that of the fairy-believer. But the intuition goes in precisely the opposite direction.

Those who acknowledge the possibility of failing to respond to moral reasons without irrationality, but not of failing to respond to epistemic reasons without irrationality, then, are still faced with a challenge to explain what makes the difference. I have suggested that there is no good theoretical reason to accept such an asymmetry. Still, I recognize the difference in intuitions, and it is troubling. Next, I will try to give an error theory for the difference in

---

[147] Rawls (1971: 432).

[148] Cf. Setiya (2007: 649).

intuitions between the two cases to address this troublingness. As we have just seen, failures to respond to non-moral practical reasons seem to intuitively go along with failures to respond to epistemic reasons rather than failures to respond to moral reasons. So we will also want our error theory to predict this.

2.5 An error theory for the asymmetry

Part of what explains why *philosophers* have thought that there is an asymmetry between moral and epistemic reasons may have been that they are in the grips of one of the views that I surveyed in the previous section. So that is a kind of explanation of why some have thought that there is a difference between the cases. But it seems a bit far-fetched to say that this is what explains the difference in intuitions and in ordinary usage. So I think we should look for a more comprehensive error theory.

Here is an attempt at such an error theory. Suppose that, as the coherentist thinks, the intuition about the meat-eater gets at something deep: that failing to respond to the normative reasons you have is not necessarily irrational, if you are nevertheless ideally coherent. Still, there are various ways in which you might be incoherent; for example, failing to intend to do what you believe you (all-things-considered) ought to do is plausibly a kind of incoherence. So, in the doxastic case, is failing to take the doxastic attitudes that you yourself judge your evidence to support.

Now, the fact that there are many different *moral* principles that different people accept is in particular very salient to us. Social disagreement about morality is widespread, and so we have no difficulty imagining someone like the meat-eater who thinks that eating meat is fine. This makes it vivid how someone can be coherent while having a moral outlook divergent from one's own. We are not likely to attribute incoherence to someone simply on the basis of

their acting out of accord with *our* moral principles. Moreover, we recognize (at least implicitly) that moral principles are normative claims. We thus see that there is room for normative contestation of these principles, and thus that there are others who will disagree and act accordingly.

By contrast, in the grass-counter and fairy-believer examples, things are not the same. It's not that we have difficulty imagining someone who counts grass, or believes in fairies. Rather, it's that we find it difficult to see exactly how the normative principles of such a person might diverge from our own in such a way as to make her ideally coherent (for example, such that she does not in any way believe out of line with her *own* evidential standards). Although this is a stipulation involved in the case, we tacitly reject the stipulation at least to some extent. While moral disagreement, and the plurality of moral standards, is very socially salient, disagreement about evidential and prudential standards is far less socially salient, so we tend to attribute our own standards to others more readily. Moreover, we are less clear and consistent in thinking of evidential and prudential standards as explicitly normative than in the case of moral standards. Many people fail to realize that they make contestable normative assumptions when they assume that prudential facts give them reasons for action or that some particular body of evidence gives them reasons for belief. This too makes the possibility of alternative normative frameworks less salient.

Our ordinary rationality-attributing practices rely on making innumerable assumptions about the background mental states of others. By holding certain mental states "fixed", as it were, we can assess the rationality of *individual* mental states, consistently with rationality being at heart a matter of coherence.[149] Not all of these background assumptions are, or can be,

---

[149] For a detailed account of the way that the semantics of rationality-talk allows us to also do this in language, see chapter 3 below.

based on specific information had about the agent in question. For example, if I tell you a story about a person who never carries an umbrella, no matter how many times he gets rained on, you are quite likely to judge that this person is irrational. In doing so, you just assume that this person prefers not to get wet when he goes outside, even though you have no *specific* information about this person whatsoever.

The suggestion that I am making is that when we are presented with an imaginary case involving a person who, by stipulation, lacks key background mental states that we tend to attribute by default, and which are therefore part of the set of background assumptions that we rely on centrally in attributions of rationality, we often tacitly reject the stipulation.[150] If we combine this with the claim from above that we are more likely to attribute our own standards to others by default in epistemic and prudential cases than in moral ones, then we arrive at an error theory that predicts the (unwarranted) difference in intuitions across the cases. This error theory predicts the plausible hypothesis that those in societies with less moral disagreement, where the plurality of possible moral standards is less salient, would be more willing to attribute irrationality to those that they take to be failing to respond to their moral reasons.

Let's now go back to the grass-counter and fairy-believer cases to see how what has been said helps to debunk those specific intuitions. Starting with the grass-counter, note that as stated so far the grass-counter example can be developed in one of two ways. On one way of developing it, the grass-counter just really is psychologically constituted such that counting grass makes her supremely happy. Desiring to be happy, she counts grass. On a second way

---

[150] This is an instance of a more general phenomenon, drawn to our attention by Tamar Gendler (2002): "where contingent correlations play a central role in our understanding and employment of a concept, and where the concept in question concerns an assessment of value, cases where we 'imagine away' the correlations may be uninformative as guides to that concept's application-conditions." In the article from which this quotation is taken, Gendler applies this lesson to the use of thought experiments concerning personal identity.

of developing it, the grass-counter is not made especially happy by counting grass, but rather desires counting grass for its own sake.

Both ways of developing the case, I think, are difficult for us to really accept. Grass-counting just seems so incredibly boring to any person like you or me that it is hard to imagine a human that is really made truly happy by grass-counting. But interestingly, try the following thought experiment: there's an *alien race* who enjoy nothing more than grass-counting. Would members of that alien race be irrational to spend their days counting grass? If your verdict now differs from the human grass-counter, this suggests that it may be your difficulty in accepting that a human could really be happy counting grass – that is, your failure to accept the stipulation of the case – that is explaining your verdict.

Equally, it seems hard to imagine someone who really accepts the normative claim that grass-counting is of intrinsic value. What on earth could motivate such a claim, especially in someone who isn't actually made especially happy by counting grass? Unlike the controversy over meat-eating, we have no familiarity with such people in everyday life.

So on both ways of developing the case, we may be tacitly rejecting the stipulations of the case. This on its own suggests that our verdicts in these cases are not really verdicts about people that are genuinely ideally coherent. But now note something further. On the first way of developing the case – where the person is made genuinely happy by grass-counting – our tendency to judge the person irrational is just as much a counterexample to a reasons-responsiveness account of rationality as it is to the coherentist view. For surely being made supremely happy by grass-counting is a pretty good reason to count grass! So, if our judgment that the grass-counter is irrational carries over to this case, and not just to the case of the person who desires grass-counting intrinsically, then again this is evidence that the intuitive

reaction is not really a mark against a coherence theory of rationality, but rather driven by a failure to imagine the case as stipulated.

In the fairy-believer case, we have little difficulty imagining someone with such beliefs, which are far more usual to encounter than the grass-counter's behavior. Nevertheless, I think we have a tendency to think that the evidence against fairies is so overwhelming, and the evidence for them so puny, that at some level the fairy-believer must really recognize that it is no good. So again, the fairy-believer fails to be ideally coherent as we imagine him. We tacitly reject the stipulation that the person is ideally coherent.

In sum, what I am suggesting is that our intuitions about rationality are in a deep way coherentist, but that we tend to attribute background mental states to agents that would make their flagrant disregard for reasons genuinely incoherent. I have been offering this as an error theory about our reactions to cases where the agent who disregards her reasons is supposed to be ideally coherent. But in real world cases, our attribution of incoherence to those who flagrantly disregard their reasons is not always a mistake, since in such cases, the believer is usually *not* ideally coherent. It is often, in fact, a reasonable assumption to think that people with beliefs that flagrantly disregard their evidence – such as someone, to take Lord's (2014) example, who thinks that the earth is less than 300 years old – do at least at some level realize that their evidence for their beliefs is no good, given the weight of the evidence establishing a history going back beyond 300 years.[151]

To elicit our intuitions, opponents of the coherentist account appeal to cases where a failure to respond to one's reasons is flagrant and obvious.[152] But for that very reason, such

---

[151] See also Davidson (2004: 192).

[152] See e.g. Lord (2014).

cases are in the real world likely to be ones where the person in question is at some level aware that he is failing to respond to his reasons, and for this reason is not ideally coherent.[153] There's a lot about our evidence that we have no choice but to know (and hence believe), and these beliefs may be jointly incoherent with crazy first-order beliefs.[154]

This explains why we tacitly reject the stipulation that the subject is ideally coherent in cases like the ones we have considered. As a rule of thumb about real world cases, that may not be so bad. It is only in relatively unusual philosophical cases – cases where those who flagrantly disregard their reasons are nevertheless ideally coherent – that the attribution of irrationality to those who flagrantly disregard prudential or epistemic reasons is an error. Often, such agents are in fact genuinely incoherent. So, although I have offered an error theory, the attribution of error need not be enormously widespread.

2.6 Modified reasons-responsiveness views?

I have argued that the asymmetry in intuitions between the practical domain (in relation to moral reasons) and the doxastic domain (in relation to epistemic reasons) is groundless, and for an error theory on which it is our verdicts about failures to respond to epistemic reasons that should be revised. The strong reasons for rejecting the thought that a failure to respond to one's moral reasons need be irrational carry over to failures to respond to one's epistemic reasons. This strengthens the case for a coherentist account of rationality.

Someone might now wonder whether a *modified* version of a reasons-responsiveness view could do an equally good, or even better, job of unifying the accounts of the doxastic and practical domains and accounting for all the intuitions. A "modified" version of the

---

[153] Ridge (2014: 238) makes a similar point.

[154] See also White (2007: 120-1).

reasons-responsiveness view, in the sense I mean to indicate, is any reasons-responsiveness view on which it is not the case that *any* failure to do that which the totality of one's reasons support counts as irrational. There are several different ways of developing this thought, however. I will take them one by one. The aim here is not just to bash some alternative views of rationality. Rather, I will suggest throughout that these views are inadequate precisely to the extent to which they depart from the coherentist account – and that the only way to make them adequate ultimately ends up collapsing into the coherentist account. So I aim to bolster the coherentist account further here.

(a) Responding to your beliefs about reasons?

One proposal might be that rationality consists in responding to your *beliefs* about your reasons. On a first interpretation of this view, rationality requires you to Φ iff you believe that you have decisive reason to Φ.[155] This claim needs a revision. As it stands, it entails that you are not rational if you believe that you have decisive reason to Φ but fail to Φ. But what if you are somehow prevented from Φ-ing? Then you may not be irrational. So we should revise the claim to refer to your *intending* to Φ: rationality requires you to intend to Φ iff you believe that you have decisive reason to Φ.

Now, the right-to-left direction says that rationality requires you to intend to Φ if you believe that you have decisive reason to Φ. As stated, it has ambiguous scope. Once it is clarified as having wide-scope (I'll defend the wide-scope view in chapter 3 below), it just is a version of the enkratic requirement, beloved of coherentists everywhere. The left-to-right

---

[155] This may be Kolodny's view. See Kolodny (2005: esp. 542).

direction of the conditional, however, adds that satisfying this requirement is *sufficient* for being rational. This seems dubious: there is no particular reason to think that the enkratic requirement is the *only* (coherence) requirement of rationality. In any case, this view is no threat to the coherentist view that all requirements of rationality are coherence requirements.

(b) Responding to your subjective reasons?

On a second interpretation, the claim is that rationality requires you to respond to your *subjective* reasons.[156] "Subjective reasons" here are supposed to be understood as somehow relative to your beliefs. On way of thinking of this is in terms of the objective reasons that one would have if one's beliefs were true.[157]

One immediate problem with this proposal is that, since the notion of a reason is a narrow-scope one (see section 0.3 above), this view appears to be committed to the problems that affect narrow-scope views of rationality (see section 3.2 below). For example, it might be true that if you are being chased by a flesh-eating monster, you have decisive reason to jump out of the window. So, if you believe that you are being chased by a flesh eating monster, you have decisive subjective reason (in the present sense) to jump out of the window. So, on the present view, you are rationally required to (intend to) jump out of the window. But surely this is not right, if your belief that you are being chased by a flesh-eating monster is utterly crazy.

---

[156] Views of this broad sort are held by Parfit (2011: ch. 5), Way (2010: see esp. 221), and Schroeder (2009).

[157] This still admits of an ambiguity. Is it enough for having subjective reason to Φ that one has a *normative* belief (for example, the belief that one has decisive reason to Φ!) such that if it were true, one would have decisive reason to Φ? Or are we tacitly restricting ourselves to *non-normative* propositions here? As I read them, Parfit (2011: ch. 5) intends the account to be restricted to non-normative propositions, whereas Schroeder (2009) does not. Though this is an important difference, it does not matter for most of the criticisms developed below. The one exception is the example given in fn. 159, which relies on interpreting the view, with Schroeder, as covering both normative and non-normative beliefs.

Surely you are at least permitted to instead give up your belief that you are being chased by a flesh-eating monster, rather than to intend to jump out of the window. Relatedly, it could easily be the case that one has some beliefs which, if true, would give one decisive objective reason to Φ, and also has other beliefs which, if true, would give one decisive objective reason not to Φ. The subjective reasons view appears to say here that one is both rationally required to Φ and rationally required not to Φ (and, oddly, that one has decisive subjective reason for both of these courses of action also). This seems implausible.[158] Worse, this situation need not be down to one's having logically inconsistent beliefs. See the footnote to this sentence for an example where it isn't.[159]

---

[158] Two possible lines of resistance. (1) One could try revising the account so one has decisive subjective reason to Φ just if, were the *totality* of one's beliefs to be true, one would have objective reason to Φ. The problem with this is that it may often be metaphysically impossible for the totality of one's beliefs to be true. In fact, the example given in the next footnote is plausibly a case of this sort. (2) One could try saying that even though each set of beliefs, if true, would give one decisive objective reason to Φ and not to Φ respectively, this fact only guarantees that one has *pro tanto* subjective reasons to Φ and not to Φ, and that one has decisive subjective reason to comply with whichever set of *pro tanto* reasons is stronger. (This may be Schroeder's view.) The problem here is twofold. First, it is very unclear how the weighting here is to be determined. If both sets of beliefs would individually guarantee *decisive* objective reason to Φ if true, it is unclear how to say which is stronger. (This is particularly evident, again, if it is metaphysically impossible for both sets of beliefs to be true simultaneously.) Second, this move creates problems for the view in explaining the "stringency" of certain apparent rational requirements. In particular, Schroeder thinks that his view can explain the intuitions behind both the enkratic requirement and the instrumental requirement. The latter explanation is supposed to work because Schroeder thinks that intending to Φ involves believing that one has weighty reason to Φ. Suppose now that one believes that Φ-ing is the necessary means to some end that one has, but believes that one has decisive reason not to Φ. On the present proposal, this will generate two competing *pro tanto* subjective reasons to Φ and not to Φ. But one of these *pro tanto* reasons must be outweighed (*pace* Schroeder, it is not enough simply to point out that each reason will be individually weighty). So it will either be rational to defy one's normative judgment or to defy one's intention. So at least one of the requirements turns out not to be stringent: one can violate it without irrationality. The coherentist view, by contrast, can preserve the stringency of the requirements by construing them with wide-scope. In this situation, one can come to satisfy both wide-scope requirements by either revising one's end or revising one's intention to Φ. The ability to deliver this sort of result was one of the fundamental motivations for Broome's original (1999, 2004) account, which favored understanding these requirements as stringent, wide-scope requirements, and not in terms of narrow-scope, *pro tanto* reasons. See section 3.2 below.

[159] Suppose that you have a sandwich in your bag and have promised to give it to your friend. Suppose you then pass a homeless man on the street. You believe that by breaking your promise and giving the sandwich to the homeless man, you can alleviate the homeless man's hunger – and that your friend does not really need the sandwich. Suppose that if these descriptive beliefs, taken together, are true, then you have decisive reason to break your promise. But suppose also that you (falsely) believe that you do not have decisive reason to break your promise: you think that you are morally obligated to keep your promise, and that this gives you decisive reason not to break it. Trivially, if this normative belief of yours were true, you would have decisive reason not

146

A second problem with the subjective reasons view is that it is not extensionally adequate in accounting for the intuitions surveyed in section 2.3. Recall the case of the meat-eater. The meat-eater might have certain descriptive beliefs such that, if those beliefs were true, he would have decisive moral reason to refrain from eating meat. For example, suppose that it is true that if meat-eating contributes significantly toward the suffering of animals, then one has decisive reason not to eat meat. (Again, if you disagree with this substantive verdict, feel free to substitute another example.) The meat-eater may agree that meat-eating does contribute significantly toward the suffering of animals. In that case, on the present definition, the meat-eater has decisive subjective reason not to eat meat. However, the meat-eater may not agree that the fact that meat-eating contributes significantly toward the suffering of animals is a decisive reason not to eat meat. As long as this is the case, the meat-eater does not seem irrational to eat meat.

Note that the subjective reasons account has the odd feature of being absolutely forgiving when it comes to a subject having crazy non-normative beliefs – she is never irrational unless her *own beliefs*, if true, would give her objective subjective reason to Φ – but being absolutely unforgiving when it comes to a subject having any false normative beliefs whatsoever – she is always irrational when her own beliefs, if true, would give her decisive objective reason to Φ, no matter how blamelessly ignorant she is of this reason-giving relation. This asymmetry is odd, and compounds the intuitive problem described with the meat-eater case above.

---

to break your promise. So, on the subjective reasons account, you are rationally required both to break your promise and not to break your promise. But this is not down to a logical inconsistency in your beliefs, since it isn't a logical truth that if your descriptive beliefs are true, then you have decisive reason to break your promise. It is simply down to your having a false normative belief about what you have decisive reason to do here.

One could try to fix these problems by stipulating that to count as irrational for not Φ-ing, the agent must actually have *normative* beliefs to the effect that she has decisive reason to Φ. So, the meat-eater, for example, must actually have the *normative* belief that he has decisive reason not to eat meat. But then we are back at the view described in part (a) above, which as I argued above is no threat to the coherentist account. The more general lesson here is that it is precisely the feature of the subjective reasons view that is non-coherentist that precludes it from accounting for the intuitions here. Someone who has beliefs that would, if true, give her objective reason to Φ, but fails to intend to Φ, does not seem incoherent if she herself doesn't recognize that these beliefs support Φ-ing. For that reason, it does not seem right to call her irrational. So the inadequacy of the subjective reasons view reinforces the intuitive power of the coherentist view.

(c)   Evidence-relative reasons?

Although the notion of "subjective reasons" was defined above as a belief-relative notion, this is not the only way in which the term "subjective reasons" is used. In fact, a binary distinction between "objective" and "subjective" sense of 'reasons' or 'ought', and the accompanying assumption of semantic ambiguity here, is quite implausible.[160] At minimum, it seems that there are intermediate notions between purely belief-relative notions and notions that are relative to all of the facts. For example, there is plausibly an evidence-relative notion.[161]

---

[160] Surprisingly many theorists take this binary view for granted. See, for example, Unger (1986: 164); Setiya (2004: 271); Gibbard (2005); Schroeder (2009).

[161] As I will make clear in chapter 3 below, I favor a more systematic form of contextualism for 'ought' on which it can take a huge variety of different bodies of information depending on context. Unlike an ambiguity view, the contextualist view attempts to unify all these usages under a single semantics. For further arguments

One might propose that this evidence-relative notion of reasons is the relevant one for a reasons-responsiveness account of rationality.[162] This might seem to nicely avoid the asymmetry between normative and non-normative ignorance that I argued afflicted the subjective reasons view in part (b) above. It is neither absolutely forgiving for ignorance of non-normative facts (when they deviate from one's evidence), nor absolutely unforgiving for ignorance of the normative facts (when one has misleading evidence about such normative facts).

There are two problems with the evidence-relative view, however. The first, and largest, is that the whole of chapter 1 above effectively constitutes an argument against the evidence-relative view. The evidence-relative version of the reasons-responsiveness view *is* the dominant version of the reasons-responsiveness view of rationality as it figures in epistemology. It is assumed that the epistemic reasons that rationality consists in responding to are those provided by one's evidence. If there are epistemic reasons outside one's evidential ken (and I suggested that this talk is not nonsensical in section 2.4 part a above), then these reasons are not amongst those that it is assumed one must respond to in order to be rational. Unfortunately, this view turns out not to effectively account for the coherence requirements associated with rational belief – in particular, (ILC) – since responsiveness to one's evidential reasons and conformity with (ILC) can come into conflict.

A second problem with the evidence-relative view is that it still does not seem to be extensionally adequate with respect to the intuitions surveyed in section 2.3. Consider once again the meat-eating case. Let us stipulate that you and I have the same exposure to evidence

that positing a small and limited number of 'oughts' is an unstable position, see Kolodny & MacFarlane (2010: 121); Pittard & Worsnip (ms.).

[162] See Kiesewetter (ms.-a), and on one interpretation Wedgwood (2003).

with respect to the effects of meat-eating, and with respect to the arguments for and against meat-eating. Still, if I conclude the meat-eating is wrong and you conclude that it is permissible, I would be inclined to say that in eating meat, you fail to respond to your (evidence-relative) reasons, but not that you are irrational. It seems here that what really explains my reluctance to ascribe irrationality to you is the fact that you are not incoherent.

(d) The reasons that you possess?

The last two versions of the modified reasons-responsiveness view have appealed to a particular *kind* of reason – your subjective (belief-relative) reasons, or your evidence-relative reasons. An alternative way of going appeals, instead, to a restricted *subset* of your total set of (objective) reasons. On such a view, rationality requires you to respond just to this subset of your reasons.

A plausible candidate for the subset, one might think, is the set of reasons that one *possesses*, or *has*, in the sense mentioned in section 2.4 part a. Errol Lord (2014b) puts forward a novel version of the reasons-responsiveness account that takes exactly this form. This view is different from the views already considered because, whereas there can be subjective reason for one to Φ without there (actually) being objective reason to Φ, one cannot have an (objective) reason to Φ without there (actually) being (objective) reason to Φ. This is what is meant by saying that the reasons one *has* are a subset of one's total reasons. This notion of one's "having" a reason to Φ, as distinct from there being a reason for one to Φ, may be partially stipulative. In fact, I think that Lord's particular gloss on what it is to have a reason is clearly not just an analysis of the ordinary locution 'have a reason'. But that is OK; we can still consider the account.

On Lord's account, two things are required to have (or possess) a reason. Suppose that some proposition R constitutes a reason for you to Φ. On Lord's view, you possess R just if (i) you stand in the right epistemic relation to R,[163] and (ii) you are disposed to *treat R as a reason* to Φ. To treat R as a reason is, roughly speaking, to make use of it in one's reasoning. Lord calls this second condition the "treating condition," and I will follow him in this usage.

It is important to see why Lord introduces this second condition (which, in my view, is what leads his account away from the most natural sense of ordinary talk of 'having' a reason). With only the first condition introduced, Lord's view would be vulnerable to the same intuitive worries raised against the views already considered in this section. For there are cases where you can be aware of the truth of some proposition R that is in fact a reason to Φ, but in no way *take* R to be a reason to Φ, and in such cases you do not seem irrational for failing to Φ. For example, I suggested, the meat-eater could be aware that meat-eating contributes significantly to the suffering of animals, and but not take this to be a (decisive) reason to refrain from eating meat. In such a case, even if the contribution to suffering is a reason not to eat meat, the meat-eater is not irrational for failing to respond to it. The treating condition solves this problem with the account. It prevents us from counting the meat-eater as irrational, because even though he is aware that meat-eating contributes significantly to the suffering of animals, he is not treating this as a reason.

I will argue, however, that the treating condition is a major concession on Lord's part to the coherentist view. Lord claims to be vindicating our ability to call cases of flagrant disregard for one's reasons "irrational," and takes this to be a major advantage for his restricted reasons-responsiveness view over the coherentist view. Yet as it turns out, Lord's introduction

---

[163] In other work (Lord 2010), he has suggested that the relation is that of being in a position to know R.

of the treating condition completely removes this purported advantage. In fact, I shall argue, Lord's restricted reasons-responsiveness account is *strictly less demanding* than the coherentist account. So my argument can be seen as a dilemma for Lord's version of the reasons-responsiveness account. Either it gives up the treating condition, and with it the ability to deal with the cases that spelled trouble for other modified reasons-responsiveness views; or it retains it, and thus loses all of its purported advantages over the coherentist view. Once again, this illustrates the general point that the modifications needed to make the reasons-responsiveness view adequately plausible are exactly those that are distinctive of the coherentist view.

I will begin by showing that, given the treating condition, every agent that Lord's view counts as irrational can be counted by the coherentist view as irrational. I then turn to showing that not every agent that the coherentist view counts as irrational can be counted by Lord's view as irrational. It follows that Lord's view is strictly less demanding than the coherentist view, and so that it cannot have the advantage Lord claims for it.

Let us have before us explicitly Lord's requirement and the coherentist's enkratic requirement, as formulated in terms of decisive reasons:

Lord's requirement. If

(i)   R is a decisive reason to Φ; *and*

(ii)   S is disposed to treat R as a decisive reason to Φ;[164] *and*

---

[164] Lord isn't totally clear that S has to be disposed to treat R as a *decisive* reason to Φ, rather than just as *some* reason to Φ. But I assume that he intends the former. In introducing the treating condition, Lord acknowledges that we shouldn't automatically count you as irrational for being mistaken about which facts support which actions. So I take it that likewise he won't want to say you're automatically irrational for getting the *weight* with which these facts support these actions wrong. So I assume that the treating condition requires that you are disposed to treat R as having its full weight.

(iii)   S stands in the right epistemic relation to R, *then*

rationality requires S to Φ.

Enkratic requirement. Rationality requires of S that (if S believes that she

has decisive reason to Φ, then S intends to Φ).

As with the view considered in part a above, one salient difference between Lord's requirement

and the enkratic requirement is that Lord's requirement pertains to whether S Φ's, whereas

the enkratic requirement pertains to whether S *intends* to Φ.[165] This may immediately suggest a

way in which Lord's requirement is more demanding than the enkratic requirement. Suppose

that you have decisive reason to Φ, and you believe this. Suppose also that you intend to Φ.

But suppose that you fail to Φ. Then, Lord's requirement, it might seem, may count you as

irrational. But the enkratic requirement will certainly not do so.

Each view faces an intuitive objection here. When we work through their natural

responses to these intuitive objections, we see that the two requirements cannot actually have

different consequences in a case like this after all.

The intuitive objection to the formulation of Lord's requirement is one that I already

raised in part a above. It is the problem that even when you believe you have decisive reason

to Φ, you may fail to Φ because you are prevented from doing so. In such a case you may not

be irrational.

---

[165] Because the enkratic requirement refers to whether S intends to Φ, it must restrict itself to treating 'Φ' as
representing potential *actions*. For it is not clear that the schematic 'intends to Φ' makes sense if 'Φ' represents a
mental state or attitude. For example, the notion of intending to believe something may not make sense. Even
if it does, it may not be plausible you are irrational to fail to have intentions to believe, as the enkratic
requirement would imply if 'Φ' could represent a belief. If we need to interpret the enkratic requirement as
referring only to actions, then it needs supplementing for the doxastic case. Fortunately, a ready supplement is
available: (ILC), as expounded and defended in chapter 1.

As I see it, Lord has two options in response. One is to concede the objection, and revise his requirement so that it requires an intention to Φ. The second is to claim that if you are prevented from Φ-ing, you do not really have decisive reason to Φ. In that case, the antecedent conditions of Lord's requirement could not be satisfied in such a case. So his requirement would not have the purported bad result. Obviously, if Lord took the first option, this would eliminate the present ground for thinking that Lord's requirement is more demanding than the enkratic requirement. So, in investigating the possible grounds for this claim, we can safely focus only on the second option.

Now for the intuitive objection to the formulation of the enkratic requirement. Suppose you believe you have decisive reason to Φ, and intend to Φ, but fail to Φ because you are just too lazy or weak-willed to do so. The enkratic requirement, it seems, fails to count you as irrational. This may seem too lenient. However, Broome, a prominent defender of the enkratic requirement, has a reply to this objection.[166] On his view, failure to Φ at the final possible moment of decision, when nothing prevents you from Φ-ing, constitutes a failure to intend to Φ. So the case as described is impossible.

Let us put each side's response to the intuitive objection to their view together. Suppose you believe you have decisive reason to Φ but do not Φ. Either you are prevented from Φ-ing or you are not. If you are prevented from Φ-ing, then you may not violate the enkratic requirement. But you will not violate Lord's requirement either, given the claim that if you are prevented from Φ-ing, then you do not have decisive reason to Φ. In any such case, the antecedent of the conditional stated in Lord's requirement will not be satisfied.

---

[166] Broome (2013: 151).

Suppose next that the case is the same, except that you are not prevented from Φ-ing. Then you will violate Lord's requirement. But likewise, given Broome's claims about intention, you will not count as intending to Φ. Consequently, you will violate the enkratic requirement.

So, whether or not you are prevented from Φ-ing, it will not end up being the case that Lord's account counts you as irrational but Broome's coherentist account does not. Consequently, the difference between the accounts in terms of referring to Φ-ing versus referring to intending to Φ turns out to be a red herring. It does not make Lord's restricted reasons-responsiveness view more demanding than the coherentist view.

Is there some other way in which Lord's requirement is more demanding than the enkratic requirement? Consider the following line of thought. The enkratic requirement says that rationality requires of you that if you believe that you ought (have most reason) to Φ, then you intend to Φ. It is thus silent in cases where you don't have any belief that you have most reason to Φ. But perhaps Lord's view is not silent in some such cases. Suppose again that R is decisive reason for you to Φ. And suppose you stand in the right epistemic relation to R for possessing R. Then, if you are disposed to treat R as a decisive reason to Φ, but fail to believe that you have decisive reason to Φ, then you might be irrational by Lord's account, without violating the enkratic requirement.

For this to happen, it would have to be possible that one can be disposed to *treat* R as a decisive reason to Φ, without *believing* it to be one. But does this case arise? That all depends on how liberally we understand belief. On some views, to believe something just is to be disposed to treat it as true; so, on this account, to believe that R is a decisive reason to Φ just is to be disposed to treat it as true that R is a reason to Φ, i.e. to treat R as a reason to Φ.[167]

---

[167] Lord makes the point that some people don't possess the concept of a reason, and thinks that such people treat things as reasons without believing them to be so. But it's far from obvious that we can't make use of

To be sure, on some accounts, it will still be possible to be disposed to treat things as true without believing them.[168] In response to this, I want to make two points. The first is that such cases are certainly in the minority. Let's remind ourselves of what Lord's broader aim was. Lord wanted to vindicate the idea that there are many cases where one is coherent but nevertheless irrational – where one simply does something that flagrantly goes against one's reasons. But once Lord refines his view to deal with the common objections to the reasons-responsiveness account, he ends up having to say that such cases are limited to ones where the subject is disposed to treat the reason as a reason. But to the extent that I share Lord's linguistic intuition that we want to describe those who flagrantly violate their reasons as irrational, I certainly don't think that we're assuming that such agents have to be ones who are disposed to *treat* those considerations as reasons without *believing* them to be reasons! For Lord's account to have an advantage over the coherentist account here, this would have to be what we're assuming about such agents when we describe them as "irrational." That is preposterous.

Likewise, consider Lord's claim to be vindicating the common wisdom of current epistemology. The common wisdom of current epistemology is that failing to believe what one's evidence supports is irrational. There is typically no condition requiring one to be disposed to treat this evidence as evidence. Lord says he wants to vindicate intuitions like the one that not believing that the earth is older than 300 years old is irrational. But it's totally

---

concepts for the purposes of belief-attribution unless those concepts are possessed by the subject of the attribution. We do this all the time with animals, young children, philosophy students, and many others besides. Moreover, in formulating his requirements, Broome (2013: 97) makes it explicit that he is using 'believes' in "a sense that is suitably transparent to the concepts through which you grasp the subject-matter of your belief." So, for example, one such requirement refers to object-given reasons, but obviously nearly no-one, in the grand scheme of things, has the concept of an object-given reason.

[168] I think the most plausible cases are those involving what Tamar Gendler (2008a, 2008b) calls "aliefs." Even these cases are far from controversial, though. On one reasonable alternative, such cases are "in-between beliefs." See Schwitzgebel (2010).

clear that someone who fails to believe this might, even if aware of the evidence against this claim, fail to be disposed to *treat* it as evidence that the earth is older than 300 years old. In fact, that seems by far the most likely thing for such a person! But since such a person fails to meet the treating condition, Lord's account does not count him as irrational any more than the coherentist account does. So I think Lord's advantage here is severely diminished.

I did promise more, however; I promised to show that the coherentist account can count as irrational *every* case that Lord's view can count as irrational. This brings me to my second response to the cases in which there's a gap between believing a proposition to be a reason and a disposition to treat it as a reason. It is a simple one: namely that the coherentist can, consistently with coherentism, revise the enkratic requirement to mimic Lord's result. She merely needs to substitute the notion of a disposition to treat for the notion of belief:

Revised Enkratic Requirement (RER). Rationality requires of you that (if you are disposed to treat R as a decisive reason to Φ, then you intend to Φ).

(RER) is still a coherence requirement, since dispositions are mental states.

The ability of the coherentist to mimic Lord's requirement shouldn't be a surprise, since when we really think about what that requirement says, it is really quite amenable to being thought of in coherentist terms. To violate Lord's requirement, one has to be disposed to treat something as a decisive reason to Φ, but fail to Φ. Essentially, then, when one violates this requirement, one fails to manifest a disposition one has in one's reasoning, or fails to act in line with one's reasoning. These sound like failures of coherence to me!

Now to be sure, Lord's requirement *also* builds in some conditions that do not refer to mental states. But these are built in as *weakening* conditions. Let me explain. Condition (i) and

(less obviously) condition (iii) of Lord's requirement are clearly conditions that state something that is not just about your mental states. But they are built into the antecedent of the requirement, and by strengthening the antecedent, they can only weaken the requirement. They weaken it because whenever their content fails to obtain, the requirement will fail to apply. Thus, ignoring the difference between intending to Φ and actually Φ-ing (which I have already argued fails to amount to a difference in strength), Lord's requirement is strictly weaker than:

> **Narrow-scope revised epistemic requirement (NRER).** If you are disposed to treat R as a decisive reason to Φ, then rationality requires that you intend to Φ.

(NRER) is just Lord's requirement with conditions (i) and (iii) removed. Since (i) and (iii) are in the antecedent of Lord's requirement, Lord's requirement is strictly weaker than (RER). But, as the name suggests (NRER) is just a narrow scope version of (RER). And as Broome has proved, narrow and wide scope formulations of synchronic requirements issue precisely the same verdicts about whether any given agent is rational or irrational at a single point in time.[169] Therefore, Lord's requirement is strictly weaker than (RER). Therefore, any violator of Lord's requirement also violates (RER), whereas the converse result does not hold. So, contrary to his initial ambitions, Lord's account does not allow us to count as irrational any cases that the coherentist account is precluded from counting as irrational.

---

[169] Broome (2013: 134, 148).

On the contrary: it appears that the cases it counts as irrational will be a strict subset of the cases the coherence account can count as irrational. For (NRER) counts as irrational *any* case where you are disposed to treat R as a decisive reason to Φ, but you do not intend to Φ. Lord's requirement will fail to count some of these cases as irrational – namely, the ones where conditions (i) and (iii) of Lord's requirement are not satisfied. For example, it will fail to count them as irrational if R is not really a reason to Φ.[170] So (NRER) is actually, it seems, *more* demanding than Lord's requirement.

This looks like a proof that Lord's view is actually strictly less demanding than the coherentist view. But Lord thinks he has a case where he can give a verdict of irrationality and the coherentist can't. And he also thinks that he can explain why any case of incoherence involves a subject in irrationality. Put together, these two claims suggest just the opposite of what we just concluded. What is going on?

Let's begin with the former claim, and Lord's case. In this case (adapted from Broome), Jane knows that a fish contains salmonella, and that salmonella causes sickness, but holds a bizarre normative theory according to which becoming sick on Tuesdays is not bad for her. None of her evidence supports this claim. But today is a Tuesday. In this case, Lord actually concedes the Jane is not irrational to eat the fish, since she is not treating the fact that it contains salmonella as a reason. But he maintains that she is irrational to *believe* that becoming sick on Tuesdays is not bad for her. So her total states are still irrational.

What entitles Lord to this claim? Lord claims that Jane is "culpable" for her belief that becoming sick on Tuesdays is not bad for her. He just says this and leaves it at that. But on closer inspection, the verdict that this belief is irrational amounts to Lord abandoning his own

---

[170] See Sylvan (ms.: 8) for a case of this sort which makes clear trouble for Lord's view.

theory. On his view, Jane is only irrational for failing to respond to reasons that she has, or possesses. And possessing a reason to Φ, on Lord's view, involves being disposed to treat it as a reason to Φ. But it seems that Jane is *not* in this case disposed to treat her evidence against the claim that becoming sick on Tuesdays is not bad for her as a reason for not believing this claim. So Lord's theory does not vindicate the claim that Jane's belief is irrational. And Lord explicitly endorses his own analysis for doxastic rationality, so it cannot be that he wants to treat the practical and doxastic cases differently here.

Of course, Lord could now just try to revise the case so that Jane is disposed to treat the evidence against the claim that becoming sick on Tuesdays is not bad for her as a reason for not believing this claim. But if he does this, then the coherentist can also allow that Jane is irrational, since she will violate (the relevant doxastic analogue of) (RER). Either way, the case fails to be one on which the restricted reasons-responsiveness account counts Jane as irrational but the coherentist theory does not.

Now turn to Lord's claim that every case of incoherence comes out as a case of irrationality on his view. Above I suggested that the coherence account can account for some failings of rationality that Lord's account can't: namely those where one falsely believes oneself to have decisive reason to Φ, but fails to intend to Φ. How could Lord resist this? His basic strategy is to claim that in such cases, the agent will always fail to respond to their reasons in *some* way. When there are incoherences, we have a guarantee that at least one of the incoherent attitudes will be a failure to respond to one's reasons. So the restricted reasons-responsiveness account can count every such agent as irrational.

But here again, Lord simply abandons his own *restricted* reasons-responsiveness theory. Suppose that S falsely believes herself to have decisive reason to Φ. And suppose further that S's evidence decisively supports not believing that she has decisive reason to Φ. S may fail to

treat this evidence as a reason not to believe that she has decisive reason to Φ. In that case, Lord's requirement does not count her belief that she has decisive reason to Φ as irrational. Suppose now that S does not intend to Φ, and consequently does not Φ. Since S does not actually have decisive reason to Φ, Lord's requirement does not count S as irrational for not Φ-ing. So Lord's requirement counts neither S's belief that she has decisive reason to Φ, not her failure to Φ, as irrational. But the coherentist account can count her as irrational, for simultaneously believing that she has most reason to Φ and failing to intend to Φ. So, contra Lord's argument, we have an instance of incoherence that the restricted reasons-responsiveness account cannot count as irrational.

That completes my defense of the claim that Lord's restricted reasons-responsiveness theory is strictly less demanding than the coherence account. Though I have discussed Lord in particular at some length, the point is not merely *ad hominem*; it illustrates something broader. The reason why Lord's theory can be no more demanding than the coherence theory is all down to the introduction of the treating condition into Lord's restricted theory. But the treating condition was introduced precisely to deal with the cases that spelled trouble for the other versions of the modified reasons-responsiveness view. The lesson that emerges from this section as a whole is this: any move that adequately deals with these problems ends up either collapsing into the coherentist account or into something that is strictly weaker than it.

That completes my defense of the coherentist account. Two related objections to the coherentist account from ordinary language may seem to remain. I have suggested that the coherentist account is best understood as claiming that the requirements of rationality are wide-scope. This means that they forbid particular combinations of attitudes. But first, in

ordinary language we often talk about the rationality or irrationality of *individual* attitudes.[171] And second, in ordinary language we often talk about the rationality of irrationality of *acts* rather than attitudes.

I accept both of these observations. The first will be defused in section 3.6, where I will explain how wide-scopers can vindicate the significance and truthfulness of our ordinary narrow-scope talk, and explicitly discussed in light of this explanation in section 3.7 part e. The second has already been discussed, in section 2.6 part d. As we saw there, Broome has a reply: if there are metaphysical connections between how one acts and what one intends, then it can be the case that given some particular act under some particular set of circumstances, one necessarily has some intention. That means that, against the background of certain other attitudes you have, we can say that your performing that act makes you irrational (or at least, indicates your irrationality).

---

[171] Lord (2014b) makes this objection.

Chapter 3
Narrow-Scoping for Wide-Scopers

3.1 Introduction

In the last chapter, I defended the view that rational requirements are coherence requirements. So far, I have been assuming that these coherence requirements are wide-scope rather than narrow-scope (I explained this distinction in section 0.3). It is now time to face this issue head on.

I need to begin by disambiguating some not always well-distinguished things that might be meant by "the wide-scope theory." Consider natural language utterances like

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

For now, let's assume that '(rationally) ought to' expresses 'rationally required to' here, though we should note that it is a desideratum of a theory that it explain how this is possible. Let's also assume that 'plan to' expresses something to do with having the relevant intention. (If you disagree, just substitute what you take to be the most natural way of talking about intentions in ordinary English into (1).)

Superficially, (1) appears to have narrow scope: 'ought' appears in its consequent, and does not seem to be taking scope over the whole conditional in which it is embedded.

However, several advocates of the wide-scope theory have claimed that this appearance is merely superficial.[172] On their view, what (1) really expresses[173] is something like

(1W) Rationality requires of you that (you intend to start keeping bees → you intend to purchase a beehive)[174]

Thus, these philosophers endorse the *semantic* thesis that ordinary conditional talk about structural requirements of rationality is to be given a wide-scope reading. It's important to see that this semantic claim is independent of accepting the truth of the wide-scope view as the correct account of the fundamental rational requirements.[175] For example, it might be the case that (1) expresses (1N) and not (1W), but also be the case that (1W) is true and even that it is (an instance of) a fundamental requirement of rationality. So when people talk about the "wide-scope theory," they need to clarify whether they are talking about a view that says that in some sense (1W) is what is meant by the *utterance* (1), or whether they are talking about a view on which (1W) is the true (and fundamental) requirement of rationality.

Nevertheless, one can see how someone drawn to the latter view might be drawn to the former. If our ordinary talk is narrow-scope, but the true requirements of rationality are wide-scope, it initially looks as though we must say that our ordinary talk about rationality is systematically false. Given that the superficial logical form of ordinary utterances can often be

---

[172] See, e.g., Dancy (1977); Broome (2013: 32, 111); Brunero (2010: 30).

[173] The claim could be, slightly more cautiously, that (1) is scope-*ambiguous* between a narrow- and wide-scope reading, but that the latter is the charitable interpretation.

[174] See Dancy (1977); Broome (2013: 32, 111); Brunero (2010: 30).

[175] See also Silk (2014: 6); Dowell (2012: 279).

misleading, it is easy to see why principles of interpretative charity might then lead us to embrace a wide-scope semantics in order to save ordinary utterances from error. The idea would be that it just isn't plausible to interpret ordinary utterances as expressing narrow rather than wide-scope requirements if only the latter hold.

Unfortunately, as we will see later, an increasing body of work suggests that the wide-scope semantics is a poor account of our ordinary talk about rationality.[176] It looks, then, like we face a dilemma. Each of the following three claims is independently appealing:

(i)   The fundamental, true requirements of rationality are wide-scope (not narrow-scope).

(ii) Ordinary talk about rationality is narrow-scope (not wide-scope).

(iii) Ordinary talk about rationality is not massively mistaken.

Yet it is initially hard to see how all of (i)-(iii) can be true. Whilst one could bite the bullet and reject (iii), it is worth seeing whether the dilemma can be avoided altogether. In this chapter I will argue that it can be: contrary to initial appearances, we can square these three claims. We can do so by implementing a contextualist version of a narrow-scope semantics which is a natural extension of an orthodox semantic treatment of modals. The basic contextualist semantics has been offered as a rival to the wide-scope semantics before.[177] What is new here is a particular proposal for implementing the contextualist view in the context of rationality-talk which makes narrow-scope sentences like (1) come out true *in virtue of* the truth of the wide-scope requirements. This preserves the explanatory fundamentality of wide-scope

---

[176] Critiques of a wide-scope semantics for modal talk go back to Hansson (1969), van Fraassen (1972), Lewis (1973) and Kratzer (1981). For more recent developments of the critique in the context of recent literature on rationality, see especially Dowell (2012) and Silk (2014).

[177] See the references in the previous footnote.

requirements while also explaining the pervasiveness – and charitably saving the truth – of narrow-scope talk. On the correct interpretation of this narrow-scope talk, I'll hold, narrow-scope claims have none of the troubling consequences that they have often been thought to have – in fact, they commit us to nothing that the wide-scope account didn't already give us. To fill out the title of this chapter, the resulting view is a form of narrow-scoping about *rationality-talk* for wide-scopers about *rationality*.

Although the view developed here is a relatively straightforward application of a standard contextualist semantics, the idea of marrying this contextualist semantics with a wide-scope account of the fundamental requirements of rationality is not one that has been pursued before. Instead, many have tended to assume that a wide-scoper about the fundamental requirements of rationality is committed to embracing the implausible wide-scope semantics. Moreover, even if one rejects that assumption, it looks like a wide-scope account *without* a contextualist semantics will commit one to attributing pervasive error to ordinary speakers. So demonstrating how to combine the wide-scope account with the contextualist semantics is a very important project in defending the wide-scope account. Moreover, I will be arguing that the combination of the wide-scope account with the contextualist semantics gives us the best of both worlds on a number of scores, and avoids or resolves several additional problems and puzzles about rational requirements.

I will proceed as follows. Section 3.2 reviews and summarizes the standard reasons for embracing the wide-scope account of the true, fundamental requirements of rationality, and the problems for a narrow-scope account of such requirements. Sections 3.3 and 3.4 explain the wide-scope and contextualist semantics for modals, respectively. Section 3.5 then compares the two, arguing for the superiority of the latter on several grounds. Section 3.6 shows how to extend the contextualist semantics to talk about rationality in a way that makes

166

the wide-scope requirements explanatorily fundamental. Section 3.7 shows how the view resolves various further puzzles about coherence requirements. Section 3.8 considers some related issues about detachment rules.

A quick point of clarification is in order before I begin. The wide-scope view of rational requirements says that the requirements of rationality are requirements to satisfy *material* conditionals. As such, the wide-scope *semantics* says that what is expressed by sentences like (1) is a requirement to satisfy a material conditional. On the contextualist semantics that I will advocate, by contrast, the conditionals in ordinary utterances like (1) are not material (or, indeed, any kind of two-place logical operator). Still, we can stipulatively use the material conditional for the purposes of stating the true rational requirements. In this chapter, when I use the symbol '→', I use it for the material conditional. When I use 'if…then', by contrast, I use it for whatever the English 'if…then' expresses in ordinary language.

Since the view I'll argue for doesn't interpret ordinary language conditionals as material, it *won't* vindicate a narrow-scoped version of (1W) that uses the material conditional. So, if one thinks of the narrow-scope semantics as interpreting (1) that way, the view I defend won't count as a narrow-scope semantics, and falls outside the wide-narrow dichotomy entirely. Nevertheless, I think there is a broader, simple sense in which the view vindicates narrow-scope talk. Utterances like (1) are narrow-scope in their syntax: the 'requires' operator occurs after the 'then' in the sentence, in what looks like the consequent of the conditional. On the wide-scope semantics, this syntax is just misleading, since such utterances really express the claim that you ought to make a material conditional true. You are not in any sense required to make the consequent true. On the contextualist account, by contrast, the narrow-scope syntax is not misleading. On this view, it can be true to say that you are required to make the

consequent of the conditional true; it's just that the relevant 'required' has to have a restricted semantic value.

3.2 Wide-scoping vs. narrow-scoping about the fundamental requirements of rationality

The basic, powerful thought behind the wide-scope view is that these coherence requirements – as structural requirements – effectively forbid certain *mismatches* of attitudes (or lack of). Let us take as our example – as we will through the whole of this chapter – the instrumental requirement, a requirement of practical rationality. The thought here is that it is irrational to intend an end E, and to believe that some means M is necessary for achieving E, but not to intend M. But that is just equivalent to saying that one must make true the conditional ((one intends end E ∧ one believes that M is a necessary means to end E) → one intends M). That is:

(IR-Wide) Rationality requires of you that ((you intend end E ∧ you believe that means M is a necessary means to E) → you intend M).

One can satisfy this requirement either by intending M, or by not intending E, or by not believing that M is a necessary means to E. The wide-scope formulation captures this "symmetry". By contrast, a narrow-scoped version of the requirement would read:

(IR-Narrow) (You intend end E ∧ you believe that means M is a necessary means to E) → rationality requires you to intend M.[178]

---

[178] Notice that I have used the material conditional here, as per my stipulation in the Introduction. As will become clear later in the chapter, I think that if the requirement were formulated with the English 'if…then', it

Here the symmetry is broken. I will briefly survey three (existing) arguments as to why this is a problem, such that we should prefer the wide-scope formulation to the narrow-scope one.

The first is simple. Suppose that you do in fact intend E and believe that M is a necessary means to E. (IR-Narrow) appears to have the consequence that in that case, you are rationally required to intend M. But it seems that this is false, or at least not necessarily true. It might be rationally permissible for you to give up your end, or even your means-ends belief, instead.[179]

The narrow-scoper might reply to this that (IR-Narrow) doesn't have this consequence. Were you to give up your end or your means-end belief, you'd be "exiting" from (IR-Narrow)'s applying to you. As such you wouldn't violate it or satisfy it. Since you wouldn't violate it, it doesn't forbid it.[180] Call this the "exiting strategy."

Whether the exiting strategy is convincing or not depends on the vexed further issue of whether (IR-Narrow) is to be understood *synchronically* – as applying at one point in time – or *diachronically* – as applying over a period of time. On the synchronic version, we get:

(IR-Narrow-Synchronic) (You intend, at t, end E ∧ you believe, at t, that M

is a necessary means to E) → rationality requires you to intend M at t.

---

actually states a *truth*, albeit only in virtue of the truth of the fundamental wide-scope requirement. The criticisms of the narrow-scope formulation in this section, however, would apply to any interpretation of the relevant conditional such that (i) it licenses *modus ponens* and (ii) 'requires' is taken to be invariant in its semantic content.

[179] See, e.g., Greenspan (1975: 272-4); Wallace (2001: 17).

[180] See Lord (2011).

It does seem correct that you could give up your end without violating (IR-Narrow-Synchronic) at any individual point. All that is needed is that you intend the means as long as you intend the end. When you give up the end, you are permitted to give up the intended means as well. So the exiting strategy appears to save (IR-Narrow-Synchronic) from the first objection.

As it happens, however, the three defenders of the narrow-scope view that I know of – Kolodny (2005), Schroeder (2004), and Lord (2011, 2014a) – all think that the instrumental requirement is to be understood diachronically. It's much trickier to figure out how to state a diachronic version of (IR-Wide) – I'll come back to this in section 3.7 part b – but that doesn't necessarily mean that (IR-Narrow) can't be understood diachronically. For our purposes here, we can just note that it's much less obvious how the exiting strategy is supposed to save a diachronic version of (IR-Narrow) from the first objection. According to a diachronic version of (IR-Narrow), when you believe that M is a necessary means to some end E that you have, rationality requires you *next* to intend M. But surely it is axiomatic that if you're required to Φ, then you're forbidden from not Φ-ing; and that if you're forbidden from not Φ-ing, then you're forbidden from doing anything that guarantees that you don't Φ.[181] So as long as the exiting strategy involves not intending M, exiting is forbidden according to a diachronic version of (IR-Narrow). The diachronic version of (IR-Narrow) thus forbids you from exiting it.[182]

---

[181] Thanks to Keith DeRose for pointing out the need for this second axiomatic claim.

[182] See also Brunero (2010: 35). Lord (2014a) responds to this sort of worry by introducing the idea that there are "cancelling conditions" which get rid of the rational requirement on you, one of which is giving up your end. I think this makes the view remarkably similar to the wide-scope one (*pace* his claim that it doesn't; see section 3.7, part c).

The two arguments that remain both apply whether we understand (IR-Narrow) synchronically or diachronically. The second argument for preferring (IR-Wide) to (IR-Narrow) concerns the threat of "bootstrapping."[183] Since (IR-Narrow) allows for "detaching", it allows that merely by having some insane or despicable end, I can just make it the case that I am rationally *required* to intend to take the means to that insane or despicable end, by a kind of act of will. That seems problematic. Maybe, if rationality is a matter of coherence, intending some despicable means is not necessarily *irrational*. But nor does it seem like it could ever be rationally *required*, or that I can make it so. True, it does seem that you would be irrational to intend the despicable end and fail to intend the despicable means. But (IR-Wide) can preserve that, without allowing that you are required to intend the despicable means *simpliciter*.

This point is especially important if one thinks that in some sense one *ought*, or has reason, to satisfy the requirements of rationality (cf. section 2.2). If that's right, I can make it the case that I *ought* to take these means, and thus to achieve the despicable end, just by intending to achieve the despicable end. I can bootstrap a reason to do something awful and worthless into existence just by intending to do it. (IR-Wide) avoids these consequences.

Third, narrow scope requirements can have some very odd consequences when different instances of them are combined.[184] Suppose that you intend to pick your son up from school, and believe that to do that you must be in your car at 3:30. But suppose you also intend to take an important call, and believe that to do that you must not be in your car at 3:30. In this case, the narrow-scoped requirement suggests that rationality requires you to intend to be in your car at 3:30 and to intend not to be in your car at 3:30. This is intuitively off. Rationality

[183] See Greenspan (1975: 263); Bratman (1987: 23-7); Wallace (2001: 16-17); Broome (2004: 30, 49); Shpall (2013: 720-1).

[184] See Gensler (1985: 156-8); Broome (2007; 2013: 136-8) and Brunero (2010), although they all make the point rather differently to how I make it here.

does not require you both to have contradictory intentions. The wide-scope view, by contrast, gets things just right here. You can satisfy both instances of the instrumental requirement by either revising your intention to pick your son up from school or revising your intention to take an important call, and then either intend to be in your car at 3:30, or intend not to be in your car at 3:30, depending on which end you keep and which you ditch.

The objections just surveyed are all ways of developing the fundamental thought that the fundamental coherence requirements of rationality are wide-scope rather than narrow-scope. Are there objections to this view? The central objection in the literature takes issue with the very feature of wide-scope requirements that I claimed as an advantage, their "symmetry." As I said, one can satisfy (IR-Wide) not only by intending M, but alternatively by giving up your end E, or even by giving up your belief that M is a necessary means to E. But some philosophers, such as Mark Schroeder and Niko Kolodny, have thought that this is the wrong result; these latter courses of action would not be a rational response to your situation, at least not under normal circumstances.[185] According to the thought here, it is good reasoning to form an intention to take a means on the basis of a pre-existing end, but it is not good reasoning to give up an end on the basis of a lack of an intention to take the means; still less to give up a belief that a means is a necessary for an end on the basis of an intention to take the end but not the means.[186, 187]

---

[185] See Schroeder (2004); Kolodny (2005).

[186] Note that if one wants to make these courses of action rationally impermissible, one must opt for a *diachronic* narrow-scope view, and not a *synchronic* narrow-scope view. As noted above, this is what both Kolodny and Schroeder both do; we can now see why.

[187] There are some rational requirements for which it's hard to see how a parallel argument would go. For example, take the requirement not to believe contradictions. The wide-scope formulation is that rationality requires of you that (if you believe p, then you don't believe not-p). But faced with a conflict between believing p and believing not-p, there is no even *prima facie* asymmetry in which belief it would in general be rational or irrational to eliminate.

As it stands, there is a sleight of hand here. As several philosophers have pointed out,[188] (IR-Wide) does not entail that it's rationally permitted to take either of these two latter courses of action. For (IR-Wide) does not claim to be the only requirement of rationality. There may be other supplementary requirements that often make these latter courses of action irrational.[189]

Still, one might claim that if taking these courses of action were *always* irrational, we have reason to reflect that in our formulation of the instrumental requirement, and to replace (IR-Wide) with (IR-Narrow). Again, however, there is a sleight of hand. Suppose – for the sake of argument – that it is always irrational to give up an end *on the basis of* a lack of an intention to take the means. Still, it surely may be the case that, faced with a conflict between an intention to achieve an end and a lack of an intention to take a means that one believes is necessary for the end, one can sometimes rationally give up the end. (For example, we don't want to say that rationality precludes one from giving up some dastardly, evil end, rather than taking the means to it.) This is compatible with the claim that it is always irrational to give up an end on the basis of a lack of an intention to take the means. For one might give up the end on a different basis (for example, a recognition of the end's badness). Yet (IR-Narrow) precludes rationally giving up the end on *any* basis in this situation. So it is not enough for the critic of (IR-Wide) to argue that it is always irrational to give up the end in situations of means-ends conflicts *on the basis of a lack of an intention to take the means*; she needs to argue that it is always irrational to give up the end in situations of means-ends conflicts *full stop*.[190] But this is

---

[188] See Way (2011); Brunero (2012); Broome (2013: 139-141).

[189] As Schroeder (2004: 346) recognizes.

[190] Kolodny tries to justify his focus on the former by proposing a "reasoning test" for determining the scope of requirements (Kolodny 2005: 518-21). For critical discussion, see Brunero (2010: 44-7); Way (2011: 235-7).

plainly false. We need the wide-scope formulation in order to allow that *sometimes*, one is permitted to give up the end.[191, 192]

Incidentally, I'm also somewhat skeptical of the claim that it's always irrational to give up an intended end even *on the basis of* a lack of an intention to take the means. Suppose I intend to go to the beach, but then I recall that the train doesn't run to the beach on Sundays, so in order to go to the beach I'd have to rent a car. I don't think there's much wrong with thinking, "well, I'm not prepared to rent a car; so, I guess I won't go to the beach after all."

As I mentioned above, the wide-scoper about the instrumental requirement can appeal to *other* requirements of rationality in explaining why it is often impermissible, when faced with a means-end conflict, to revise an end or to give up a belief that some means is necessary for that end. But it's also worth noting that it may be that it is simply one's *reasons* that determine how one should respond in some situations where one violates the instrumental requirement. If it often seems like giving up one's end, or in particular giving up one's belief that the means are necessary for the end, is the "wrong" response to the situation, that may be because the *reasons* don't support doing so. Conversely, sometimes the reasons support giving up one's intention, as with the evil intentions. But failing to respond to one's reasons is not necessarily a failure of rationality, as I argued in chapters 1 and 2. So it's not obvious that we actually need to design new requirements of rationality to account for these failures.

---

[191] See Brunero (2010: 41); see also Korsgaard (2008: 49); Scanlon (2004: 234-5).

[192] What about the means-ends belief? Interestingly, Jonathan Way (2010: 223) prefers a version of the instrumental requirement that is a kind of scope-hybrid, with the means-ends belief outside the scope of the deontic operator, but the intended end inside its scope: "If you believe that doing M is necessary for doing E, you are rationally required to [intend to do M, if you intend to do E.]" However, I don't think this differential treatment is justified. There are easily envisageable situations where you can rationally give up the means-ends belief too; see Brunero (2012: esp. 133).

That said, where the reasons are just obvious, such that you can't help but acknowledge them, these revisions will actually mean that you count as irrational, due to rational requirements that we have already encountered.[193] For example, suppose you intend to visit your sick grandmother because you realize that you ought to do so. However, you then realize that in order to visit your sick grandmother, you must get your car fixed, and you don't intend to do that, because you're lazy. If you give up your intention to visit your sick grandmother, then you will violate the enkratic requirement, since you believe you ought to visit her.[194] Moreover, it's likely that you realize in a simple case like this that your evidence supports believing that in order to visit your sick grandmother, you must get your car fixed. So, if you give up your belief that in order to visit your sick grandmother, you must get your car fixed, you will violate Inter-Level Coherence. That leaves forming the intention to get your car fixed as the only option that won't put you in violation of a requirement.[195]

In conclusion, I think that wide-scoping provides the correct account of the nature of the fundamental coherence requirements of rationality.[196] But what about our *talk* about rationality? We turn to this next.

3.3 Wide-scope semantics

In understanding how someone might advocate a wide-scope semantics for ordinary claims, it's helpful to begin – as proponents of this view often do – with modal claims. Indeed, let's

---

[193] See also Scanlon (2007: 94); Shpall (2013: 722).

[194] And getting rid of *that* belief may not be possible for you.

[195] Putting all of this together, I think we may be able to do without "basing prohibitions" as supplementary requirements of rationality (cf. Way 2011; Broome 2013: 140-1, 186-91). But I'll stay officially neutral on this.

[196] I'll consider a different objection to the wide-scope account with my semantics in hand in section 3.7, part c.

initially start with non-deontic modals, and with necessity-claims. Consider the intuitively reasonable utterance

(2) If there are at least three apples, then there must be at least two apples

The naïve narrow-scope reading of (2) would be:

(2N) There are at least three apples → □ (There are at least two apples)

But (2N) is just patently false. Narrow-scope material conditionals allow us to "detach" their consequents via *modus ponens*. So if (2N) were true, then the following argument would be sound:

P1. There are at least three apples.

P2. There are at least three apples → □ (There are at least two apples)

C. □ (There are at least two apples)

But this argument does not look sound. P1 tells us that there at least three apples, but it makes no claim to being a necessary truth. By contrast, C says that it is a necessary truth that there are at least two apples. But it does not follow from its being a contingent truth that there are at least three apples that it is a necessary truth that there are at least two apples.

Consequently, say advocates of the wide-scope semantics, (2N) is a non-starter as a reading of (2). But (2) does seem like a reasonable utterance, and one that strikes us as true. So perhaps we should read the 'must' in (2) as taking scope over the whole conditional:

(2W) □ (There are at least three apples → There are at least two apples)

(2W) is true: it is a necessary truth that if there are at least three apples, there are at least two apples. And unlike (2N), it doesn't allow us to "detach" the consequent of the embedded conditional to get the false result that necessarily, there are at least two apples. So (the argument goes), given that (2) doesn't strike us as false, (2) must really be expressing or at least "getting at" (2W).

If this argument succeeds, it shows that ordinary language is somewhat misleading or imprecise when it comes to scope. What look like narrow-scope claims are often intended as wide-scope. The wide-scope semantics has a kind of explanation of this, which is that English does not have a comfortable way of putting modals unambiguously outside the scope of conditionals: there is no easy way to express (2W) in natural language with the modal 'must'. The closest we can get is 'it must be that, if there are at least three apples, then there are at least two apples', but this is at least somewhat awkward.

If all this is right, it wouldn't be surprising if similar observations held for deontic modals, such as deontic usages of 'ought'. Consider

(3) If you are going to punch your brother, then you ought to punch him softly[197]

On a naïve narrow-scope reading, (3) is interpreted as

---

[197] The example is a variant of the so-called "paradox of gentle murder", which was introduced by Forrester (1984).

(3N) You are going to punch your brother → You ought (you punch your brother softly)

Again, this licenses detaching:

P1. You are going to punch your brother.

P2. If you are going to punch your brother → You ought (you punch your brother softly)

C. You ought (you punch your brother softly)

Again, the argument does not seem sound. You ought *not* punch your brother, either softly or otherwise, and you can't just make it the case that you ought to do so by making it the case that you are going to punch him – this seems like a form of "bootstrapping" like that discussed in section 3.2. The wide-scope semantics can give the pleasing diagnosis that this argument fails in just the same way that the argument for the necessity of there being at least two apples fails. It can then propose reading (3) as

(3W) You ought (you are going to punch your brother → you punch him softly)

Again, this blocks the detaching maneuver. Moreover, in this case the wide-scope explanation of why we tend to express (3W) with (3) is even better. For there is no way of expressing (3W) accurately and grammatically in English. Whereas one can express the consequent of P2 as "you ought to punch your brother softly", one cannot attach "ought to" to a full conditional.

From here, it is a small step to the diagnosis of (1) as expressing (1W). Assuming that 'rationally ought' expresses a rational requirement, we just have to interpret the modal 'ought' as scoping over the whole conditional in (1) just as we interpret the modal 'ought' as scoping

over the whole conditional in (3). Again, this sort of maneuver nicely avoids the threat of "bootstrapping". Although the end in (1) is innocuous, we can substitute in some insane or despicable end in its place. It does not seem that merely by having some insane or despicable end, I can just make it the case that I am rationally *required* to intend to take the means to that end, by a kind of act of will. Yet a naïve rendering of the narrow-scope claim appears to have this consequence, whereas the wide-scope interpretation avoids it.

3.4 An alternative: contextualism

Although this argument is initially tempting, I think it is far from conclusive. In this section, I will explain how a contextualist semantics for modals provides an alternative account that solves the problems with the naïve narrow-scope semantics.

The contextualist semantics for modals is actually the orthodox view in linguistics and philosophy of language, following the work of Angelika Kratzer.[198] On the contextualist view, the semantic value of modals like 'ought' varies according to conversational context. On Kratzer's picture specifically, there are two parameters in the semantics for modals that get fixed by context: the "modal base" and the "ordering source." The modal base is a set of background information that is being held fixed, and it determines a set of possible worlds that we can call "live".[199] The live worlds are those compatible with the modal base: so, if a proposition p is in the modal base, then all the live worlds will be worlds in which p is true.

---

[198] Kratzer (1981) is the canonical source for the view I am drawing on. Kratzer (2012) contains a revised version of this paper along with other important papers of hers on related matters. Kratzer's view also draws significantly on, and shares similarities with, work by David Lewis (1973). For more recent developments of the contextualist view, see, e.g. Björnsson & Finlay (2010); Cariani, Kaufmann & Kaufmann (2013); Dowell (2012); Wedgwood (2006: esp. 151-9).

[199] Kratzer calls them the "accessible" worlds.

Modals quantify over these worlds. The ordering source is what then ranks the live worlds. In the deontic context we can think of the ordering source as a salient set of norms,[200] so that the worlds get ranked according to whether those norms get satisfied. 'S ought to Φ' is then true iff S Φ's in every world at the top of the ranking – in other words, if S's Φ-ing is necessary, given which worlds are live, for the norms to be satisfied (to the greatest extent that they can be[201]). 'S may Φ' is true iff S Φ's in some of the worlds at the top of the ranking – in other words, if S's Φ-ing is compatible, given which worlds are live, with the norms being satisfied (to the greatest extent that they can be).[202]

The final piece of the puzzle that we need is a basic story about how conditionals and modals interact on the contextualist picture. The crucial idea is that the antecedent of a conditional acts as a "restrictor" which affects the semantic value of any modals that occur in the consequent.[203] On the Kratzerian framework, we can think of the proposition that the antecedent expresses as getting temporarily added to the modal base for the purposes of evaluating the modal in the consequent, thus restricting the domain of worlds over which it

---

[200] I don't mean this to sound too much just like any old social norms will do. The salient norms in a context may be, for example, the *true* moral norms. Speakers don't just get to speak truly about what one morally ought to do by making their own moral norms salient. See also Dowell (2012: 283).

[201] I prefer to avoid the term 'best worlds', since it may suggest a kind of consequentialism in the ordering source that we need not be committed to; see also Wedgwood (2006: 149).

[202] Needless to say, the exact picture just sketched is the subject of controversy, and some recent contextualist accounts of 'ought' depart from Kratzer's framework in important ways. See, e.g. Finlay (2010); Björnsson & Finlay (2010). But I think that an account that utilizes at least the basic Kratzerian framework will be partially justified by the explanatory power that we'll see it to have shortly. Of course, that does not mean that we must accept every detail of Kratzer's view; many semanticists want to tinker with it in one way or another, and I myself do not agree with everything that Kratzer says.

[203] So, on this view, the ordinary 'if…then' does not express the material conditional (cf. section 3.1 above), or indeed pick out any *operator*.

quantifies to the worlds in which the proposition expressed in the antecedent is true. The conditional will be true just if given this restriction, the consequent comes out true.[204]

Let's now see how this handles the examples that we considered in the previous section, beginning with the example featuring 'must':

(2) If there are at least three apples, then there must be at least two apples

On the contextualist view, we understand the 'must' in the consequent of (2) as quantifying only over the restricted domain of worlds in which the antecedent is true – i.e., the worlds in which there are at least three apples. Given that restriction, 'there must be at least two apples' comes out true, since it really is the case that in *all* of the contextually salient worlds (those in which there are at least three apples), there are at least two apples. So the conditional comes out true. So here we have an explanation of why (2) is true that doesn't require us to reinterpret the modal as taking wide scope over the whole conditional. Rather, it takes scope only over the consequent, as it appears to, but quantifies only over the worlds compatible with the antecedent.

Where does this leave us with detaching? Given the truth of (2), are we licensed to detach 'there must be at least two apples' when we have the premise that there are at least three apples? The short answer is that one can detach 'there must be at least two apples', but

---

[204] This picture is slightly complicated in Kratzer's later work by so-called "double modalization" readings of conditional modal claims, whereby the overt modal is interpreted as being within the scope of an additional, covert epistemic modal. See Kratzer 2012: 106-7. I'll ignore this complication. Kratzer does not think that these double modalization readings should be the *only* readings of conditional modal claims, except in special cases. Dowell (2012: see esp. 283-5, 293) gives an account of how to achieve desirable contextualist results on an account where conditional modal claims are generally interpreted as doubly modalized. I think one can also get the results she wants with the simpler account, but cannot pursue this here. I work with the simpler account here simply for the sake of ease.

only if one holds the semantic value of 'must' constant, as quantifying only over the worlds in which there are at least two apples.[205] However, there may be a risk in repeating the sentence 'there must be at least two apples' in other contexts.[206] On the contextualist view, the sentence 'there must be at least two apples' can express different propositions depending on context. For example, 'must' can have an unrestricted semantic value whereby it quantifies over all the possible worlds, or it can have a restricted semantic value whereby it quantifies over the worlds in which there are at least three apples. The sentence expresses a false proposition on the first semantic value, and a true proposition on the second. What we have to be wary of, then, is uttering the sentence 'there must be at least two apples' in a context where it will express the unrestricted, false proposition.

We can use a similar explanation for

(3) If you are going to punch your brother, then you ought to punch him softly

---

[205] See also Silk (2014: 7-8); Dowell (2012: 286).

[206] This notion of detachment assumes that it is something one does with a *sentence*, involving a kind of utterance. But it's natural to think that there is another notion of detachment on which it is something one does with *propositions*, involving a logical operation or deduction. However, on this notion of detachment, it's not clear that the question of whether one can detach the consequent of an ordinary language conditionalized modal claim really even makes sense on the Kratzerian semantics. On the Kratzerian semantics, the ordinary language conditional is a restrictor, not a two-place operator that takes two propositions as its arguments. Thus, on this view, there is no intelligible question of whether one can detach the "proposition in the consequent" of the ordinary language (as opposed to the stipulatively material) conditional. There is only the question of whether one can truly utter the sentence that follows the restrictor clause, omitting the explicit restrictor (and the answer is: it depends on context).
　　This bears on the vexed question of whether Kratzerian contextualists need deny *modus ponens* (cf. Finlay (2010: 82ff); Kolodny & MacFarlane 2010). One would think that *modus ponens* is a logical rule that relates propositions, not sentences. (After all, if we have an argument which looks superficially like an instance of *modus ponens*, but turns out to rest on an equivocation between two different uses of a word, we do not take the invalidity of such an argument to be a counterexample to *modus ponens*.) Now obviously, a contextualist does not have to deny the validity of *modus ponens* for the material conditional. When it comes to the ordinary language conditional, however, it's again not obvious that the question of whether *modus ponens* holds for it makes sense, as long as we are understanding *modus ponens* as a logical rule that relates propositions. The contextualist does need to say that utterances of the sentences that follow restrictor-clauses may come out false when those restrictor-clauses are omitted. But it's not obvious to me that this counts as a failure of any putative logical rule any more than the equivocation cases do.

The antecedent restricts us to worlds in which you are going to punch your brother. Once we restrict ourselves to these worlds, the worlds that come out ranked highest by the ordering source (the salient norms) are the ones in which you punch him softly. However, 'you ought to punch him softly' will express a false proposition if 'ought' is left unrestricted, since the highest-ranked worlds out of *those* worlds are the ones in which you don't punch him at all.

It's crucial not to misunderstand what the contextualist theory says here. What restricts us to the worlds in which you are going to punch your brother is not the *fact* that you are going to punch your brother. Rather, it is the *utterance* of 'if you are going to punch your brother.'[207] This adds the proposition that you are going to punch your brother to the modal base for the purpose of determining the semantic value of 'ought'. Moreover, the mere fact that you are going to punch your brother does not force us to hold this fact constant (by adding it to the modal base) for the purpose of making claims about what you ought to do. Even if you are going to punch your brother, we can still exclude this fact from the modal base, and truly utter sentences like 'you ought not punch your brother softly, because you ought not punch him at all.'[208] Whether this proposition gets added to the modal base (for utterances that are not explicitly restricted by it) is at least partly up to us as speakers, and depends on our conversational purposes and interests.[209] If our purposes are critical, there may be specific

---

[207] That's not to say that this is the only way a proposition can get into the modal base is by being uttered, but it is what gets this particular proposition into the modal base.

[208] Even if this fact is in the modal base, there may be another explanation of why this claim can be true, which appeals to the distinction between the "strict 'ought'" (which requires *perfect* satisfaction of the requirements that order the worlds) and the "'ought' of the best-case" (which requires the *best* available satisfaction of these requirements). I hope to pursue this idea, and its integration into the contextualist framework, elsewhere.

[209] I think this holds even if we know the proposition. This is an important feature of the contextualist account that I favor. Again, I think there are independent grounds to think this. I develop this point with respect to epistemic modals in Worsnip (forthcoming-a).

reasons for excluding propositions like that you are going to punch your brother.[210] We may not want to treat this fact as fixed for the purposes of saying what you ought to do in the sense we're interested in.

This bears on the issue of bootstrapping. Recall that the objection there was that it shouldn't be possible for you to make it the case that you ought to punch your brother softly, just by making it the case that you are going to punch him. But with the clarification just made in place, we can see that this is no commitment of the contextualist view. Let $ought_U$ be the unrestricted 'ought', quantifying over all the possible worlds. And let $ought_R$ be the restricted 'ought', quantifying over just the possible worlds in which you punch your brother. Then there are two propositions: that you $ought_U$ to punch your brother softly, and that you $ought_R$ to punch your brother softly. The first is false and the second is true, but neither is made true by your deciding to punch your brother. Both are true independently of that decision. Nor is it even true that by deciding to punch your brother, you somehow make any subsequent utterance involving 'ought' pick out $ought_R$.

So there is no bootstrapping here. It's not that in the worlds in which you're going to punch your brother, you $ought_U$ to punch him softly. That would be like saying that in the worlds in which there are at least three apples, it's an (unrestricted) necessary truth that there are at least two apples. That is false, and it's not the way restriction works in the contextualist account.

The point here is parallel to one about a common confusion about contextualism about 'knows'. Contextualists about 'knows' posit that the semantic value of 'knows' varies

---

[210] I'm inclined to think we can *reject* proposals to add them to the modal base, as a kind of refusal to treat them as fixed.

depending on conversational context of the speaker: it expresses a more demanding concept in high-standards contexts. Often, this view gets misunderstood, and someone says that the contextualist says that entering a high-standards context can cause you to *lose* knowledge.[211] But this is completely wrong.[212] Contextualism says that the word 'knows' has a different semantic value when *uttered* in different contexts. By the contextualist's lights, it is still the case that S knows$_{LO}$ ordinary propositions in the high-standards context, and that S does not know$_{HI}$ ordinary propositions in the low-standards context. It's just that, in each case, the context makes it the case that the proposition in question cannot be expressed using the bare word 'knows.' So, there is no individual semantic value for 'knowledge' on which it can be said that the subject 'loses' knowledge in high-standards contexts. The mistaken idea that contextualists about 'knows' have to say that you make it the case that you lose knowledge by walking into a philosophy seminar is just parallel to the mistaken idea that contextualists about 'ought' have to say that you make it the case that you ought to punch your brother softly by making it the case that you are going to punch him.

3.5 Contextualism vs. wide-scope semantics

As we have now seen, both the wide-scope semantics and the contextualist semantics can account for the truth of utterances like (2) and (3) without the obviously bad results of an invariantist narrow-scope semantics. I'll now argue that the contextualist semantics is superior to the wide-scope semantics. I will focus on points of comparison that relate to the sorts of issues that we have been discussing in this chapter. But I do want to mention at the outset that

---

[211] Cf., e.g., Dretske (2005: 18-19). David Lewis, a prominent contextualist about 'knows', encourages this confusion in his (1996).

[212] See DeRose (2009: ch. 6, esp. 212-225).

perhaps the most major motivation for a contextualist semantics for deontic modals has relatively little to do with these points. It relates to the apparent information-sensitivity of 'ought'. Sometimes, when we use 'ought', we seem to be using it to talk about what you ought to do relative to all the facts; other times, to talk about what you ought to do relative to your beliefs; still other times, relative to some particular restricted subset of the facts such as the knowledge or evidence that you possess. One might hope that this phenomenon too can be explained by the wide-scope semantics. But in Pittard & Worsnip (ms.) I argue that it cannot be explained without contextualism, based on a systematic comparison of how the two views (and others) handle information-sensitivity. So there are independent grounds for preferring contextualism to the wide-scope semantics.

That said, let's focus more narrowly on the kinds of issues that have occupied us in this chapter so far. Begin by considering the following slight modification of (2):

(4) There are at least three apples. So, there must be at least two apples.

In broadly the same way that (2) looked like a reasonable claim, so does (4). (Of course, it may not be true if there aren't at least three apples – but it's a good inference.) So our semantics should be able to explain its reasonableness. Begin with the contextualist account. The contextualist can say that when the first sentence is uttered, the speaker proposes to add the proposition it expresses to the modal base.[213] Given that, the explanation of (4) runs in just the same way as the explanation of (2): the first sentence restricts the domain for the

---

[213] The idea of utterances as proposals to add propositions to the modal base (or "common ground") is familiar from the work of Stalnaker (1999: chs. 1-5; 2002) as well as from the "update semantics" of Veltman (1996) and Gillies (2001).

occurrence of 'must' in the second, so that the second sentence comes out true, provided that the first sentence is.

By contrast, the wide-scope semantics cannot explain the truth of (4). There isn't any conditional for 'must' to take wide scope over here. Clearly, (4) does not just express (2W), since (4) asserts that there actually are at least three apples, and that there actually are at least two apples, which (2W) doesn't. So it seems that the wide-scope interpretation has to claim that the semantic content of (4) is (i) that there are at least three apples, (ii) that necessarily, if there are at least three apples, there are at least two apples, and (iii) that there are at least two apples. But this doesn't seem like a very plausible *semantic* interpretation of (4). It says that the 'must' in the second sentence of the utterance takes scope over a hidden, unarticulated conditional that is still somehow part of the semantic content of (4). This is very cumbersome. The contextualist semantics has a significant advantage over the wide-scope semantics here.

Of course, the wide-scope semantics was primarily meant to explain conditional utterances, and (4) is not a conditional utterance. But the similarities between (4) and (2) suggest that they should be explained in a similar fashion, and so it is a disadvantage of the wide-scope semantics (as opposed to the contextualist account) that it cannot generalize beyond conditionals. If one has to call on contextualism anyway to explain the truth of (4), then the contextualism can explain (2) as well, without the need for the wide-scope semantics.

The point generalizes – we could construct a parallel argument for a modified version of (3). More generally, contextualism has the resources to explain a much wider variety of speech than the wide-scope account, since lots of speech involving modals doesn't obviously feature any conditionals. We can talk about what must be true, or what one ought to do, holding in the background various assumptions that restrict the domain that we quantify over.

For another illustration of a similar point, we can actually turn back to an example we've already encountered, namely:

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

I've deliberately avoided spelling out the contextualist semantics for uses of 'rational' and 'rationally', because I'm coming to that in greater detail in section 3.6. But we can make an observation about a limitation of the wide-scope semantics here. On the wide-scope semantics, (1) actually expresses

(1W) Rationality requires of you that (you intend to start keeping bees → you intend to purchase a beehive)

But the problem with this interpretation is that, as stated, (1W) actually isn't true, even by the lights of prominent advocates of wide-scope requirements and of wide-scope semantics.[214] At minimum, it is missing a clause in the antecedent of the conditional, which will read something like (there's room for dispute about the details): you believe that purchasing a beehive is a necessary means to starting to keeping bees. After all, you are not irrational for failing to intend something that you don't realize is a necessary means for executing some other intention of yours. So the true wide-scope requirement is really something more like:

---

[214] See Broome (2013: 159-63).

(1W*) Rationality requires of you that ((you intend to start keeping bees ∧ you

believe that purchasing a beehive is a necessary means to starting to keep bees)

→ you intend to purchase a beehive)

(1W*) pretty obviously isn't the semantic content of (1). Yet nevertheless, (1) sounds perfectly reasonable. Why is that? If (1) sounded reasonable to you when you read it first, that's probably because you were just taking for granted that the 'you' it refers to is a person who knows that one needs to purchase a beehive to start keeping bees. That's a reasonable assumption, since this is a pretty widespread piece of knowledge. Still, that assumption doesn't render (1W) true. So it doesn't help the wide-scope semantics explain how (1) is true.

But it can help the contextualist semantics explain how (1) is true. Since we're all assuming that you know that one needs to purchase a beehive to start keeping bees, it's reasonable for us to assume that this proposition is in the modal base for the utterance of (1) without it having to be explicitly stated in the antecedent. So, (1) can come out true. (More on exactly how the contextualist semantics for talk about rational requirements goes in the next section.)

Again, the point is very general. I might say that if you believe that it's raining outside, then rationally you ought to carry an umbrella. In saying this I just take it for granted that you prefer not to get wet. The contextualist account handles this by saying that the proposition that you have this preference is in the modal base, and so I speak truly. Again, the wide-scope semantics has to say by contrast that strictly speaking what I say is, as it stands, false.

Moreover, there is another major challenge for the wide-scope semantics to meet. Many ordinary claims about rationality – utterances like (1) – use 'ought'. So far we have just been granting that '(rationally) ought to' can express the same thing as 'rationally required to'.

But as noted in section 0.6, many philosophers think that 'ought' is systematically connected to *reasons*, rather than to rational requirements. So the wide-scope semantics owes us a story about why 'ought to' can express the same thing as 'rationally required to', or at least an explanation of why the latter entails the former such that (1) comes out true in virtue of rational requirements.

One simple proposal might offer itself up here: it's just the case that you ought, in the so-called "unqualified" sense of 'ought' that is connected with what you have most reason to do, to satisfy the rational requirements. In that case, we can actually just formulate the rational requirements using an unqualified, all-things-considered 'ought'. For example, we might write:

(Ought-IR-Wide) You ought (i.e. have most reason) to make it the case that

((you intend end E ∧ you believe that M is a necessary means to E) → you intend M).

Indeed, in his early work Broome used to state rational requirements using the unqualified 'ought' – a practice he shares with early forerunners of the wide-scope view.[215] The rationale running through several of these works seems to be something like the following. Given that the conditional in (Ought-IR-Wide) is material, and assuming that one can substitute logically equivalent propositions within the scope of 'ought', one can actually rewrite (Ought-IR-Wide) as:

---

[215] Broome (1999, 2004); Hill (1973); Greenspan (1975); Dancy (1977); Darwall (1983); Gensler (1985). See also Wallace (2001: 17).

(Ought-IR-Wide-Rewritten) You ought (i.e. have most reason) to make it
the case that ((¬you intend end E) ∨ (¬you believe that M is a necessary means
to E) ∨ you intend M).

Suppose that you justifiably believe that M is a necessary means to E. The idea, now, is that
when E is an impermissible end, you ought to give it up; when E is a required end, you ought
to intend M, and when E is a permissible but not required end you may either intend M or
give up E, but ought to do one of the two. (And if your belief that M-ing is a necessary means
to E is unjustified, you should give it up). In all circumstances, however, it'll be the case that
you ought to satisfy at least one of the disjuncts of (Ought-IR-Wide-Rewritten) – it's just that
further facts may determine *which*. In this way, even if our account of rationality is coherentist,
it looks like we might be able to preserve the thought that you ought (in the unqualified sense
that expresses what you have most reason to do) to be such that you satisfy the requirements
of rationality.

Unfortunately, however, there are various reasons to worry about (Ought-IR-Wide)
and analogous versions of other rational requirements. For example, Kieran Setiya has argued
that in some situation where you find yourself unable to give up some bad intention, it may
be that what you ought to do all-things-considered is violate the instrumental requirement.
Since you are unable to give up the intention, failing to intend the end is off the table; given
that, it may be that you should not intend the means even though you do intend the end.[216]

---

[216] See Setiya (2007); see also Greenspan (1975); Schroeder (2009: 227). Setiya himself takes his point to be an
objection to the idea that wide-scoping solves the bootstrapping problem that applies to the narrow-scope
version of the instrumental requirement. But this way of seeing things itself depends upon assuming that the
relevant deontic notion governing the wide-scope requirement is the all-things-considered 'ought'. Things seem
different when we talk of what is *rational*. If your end is unrevisable, then it *does* seem true that the only way to
be *rational* is to intend the means. Perhaps it is the case that you still ought not take the means – if so, what we

More simply, it seems that for any rational requirement we can dream up some scenario where your practical reasons support violating it – all we need is to have some maniac threaten to kill your family unless you do so. If you are able to violate the requirement under these circumstances, you ought to. Finally, the argument of chapter 1 has given us strong independent reason to doubt the coincidence of what one is rationally required to do on one hand, and what one has most reason to do – i.e., what one unqualifiedly ought to do – on the other. So overall, it looks bad for this claim of coincidence. If that's right, then the wide-scope semanticist cannot use that putative coincidence as an explanation of why rational requirements can be expressed with 'ought'.

Even if the two did coincide, however, there is also a more complex reason to be unsatisfied with this explanation of ordinary utterances.[217] Consider utterances that are like (1) but which involve dastardly ends, such as

(5) If you plan to train up a swarm of killer bees to kill your enemies, then you ought to plan to purchase a beehive

On the present proposal, what explains the truth of (5) is the fact that it expresses

(5W) You ought (i.e. have most reason) to make it the case that ((you intend to train up a swarm of killer bees to kill your enemies ∧ you believe that buying a beehive is a necessary

[217] Silk (2014: 4-5) makes a similar argument to the one I'm about to make. See also Greenspan (1975: 261-2); Kratzer (2012: 67); Dowell (2012: 274).

means to training up a swarm of killer bees to kill your enemies) → you intend to buy a

beehive)

Given that the conditional is material,[218] this follows trivially from the fact that you ought not

to train up a swarm of killer bees to kill your enemies. This is an instance of the general strategy,

which we are presently considering, of showing why you in general ought to satisfy the wide-

scope norms purely in virtue of the substantive, particular reasons that you have in the

individual case. Either you will have most reason to take the means or to give up the end; in

this case, you have most reason to give up the end.

But notice that, if that is what explains the truth of (5), it should also explain the truth

of the following utterances:

(6) If you plan to train up a swarm of killer bees to kill your enemies, then you (rationally)
ought to plan to purchase a 1969 Joe Morgan baseball card

(7) If you plan to train up a swarm of killer bees to kill your enemies, then you (rationally)
ought to plan *not* to purchase a beehive

After all, as long as you ought not to train up a swarm of killer bees to kill your enemies, it will

follow that you ought to make true any disjunction involving not training up a swarm of killer

bees to kill your enemies. But we do not hear (6) and (7) and true in the same way that we hear

(5) as true. This suggests that the mere fact that you ought to give up a end is not enough to

---

[218] The materiality of the conditional might in and of itself be thought a good reason to doubt the wide-scope semantics, since the English 'if…then' does not normally express a material conditional. See Dowell (2012); but see also Silk (2014: 3).

make a claim of the form 'If you intend the end, then you ought to intend the means' sound good to us. That casts strong doubt on whether the current explanation of why one can use 'ought' in utterances like (5) – the explanation that appeals to the fact that one always has *most reason* to satisfy the material conditional embedded in (5W), one way or another – is correct.

The more general point here is that we seem now to have lost the guiding thought that made the view that the true rational requirements are wide-scope so plausible: that there is something distinctively irrational in *combining* an intention to take an end, a belief that a means is necessary for that end, and a lack of an intention to take the means.[219] This is not just something that falls out of your reasons to take various ends and means; it is a *sui generis* sort of requirement.[220] Yet we want, compatibly with that, to be able to explain why we can use 'ought' to express it. That is something that the wide-scope semantics is unable to do.

Taken together, I think these points show that the wide-scope semantics does poorly in explaining ordinary utterances involving the interaction of modals and conditionals. However, this does not mean that wide-scope requirements are not true. Just to reinforce this point, let me exploit the analogy to the necessity case one last time. Recall (2W):

(2W) □ (There are at least three apples → There are at least two apples)

---

[219] In his first article on the subject, Broome (1999: 400, 402) tries to capture this by saying that the conditional involved is the material conditional plus "determination", where that means that in some way there has to be some relation between the attitude in the antecedent and that in the consequent. As far as I can tell, he drops this practice later. If one gives up the view that the conditional is material, one loses the explanation sketched above of why one always (unqualifiedly) ought comply with the instrumental requirement. So this just underscores the incompatibility of that explanation with strategies to defuse problems of triviality.

[220] One might worry about whether a similar objection trading on trivial instances of the wide-scope material conditional can be generated just for the view that there are wide-scope requirements (irrespective of our semantics). I'll address that in section 3.7, part c, below.

Regardless of whether it is the semantic content of (2), (2W) is *undeniably* true. The failure of the wide-scope semantic theory is no mark against (2W)'s truth in the slightest. Indeed, I now turn to showing how the wide-scope requirements of rationality can actually *explain* the truth of (contextually variable) narrow-scope talk.

3.6 Extending contextualism to (talk about) rationality

Someone might wonder whether the contextualist story for modals transfers over to talk about rational requirements. For quite some time, I thought that it did not. As I noted above, one striking feature of the context-sensitivity of 'ought' is the way that it is information-sensitive. We can talk about what you ought to do or believe relative to all the facts, or relative to some subset of them, or relative to your beliefs, and so on.

One thing that allows for this information-sensitivity, it seems, is the close connection between (at least many usages of) 'ought' and reasons. Adding or taking away information seems to make a difference to which reasons factor into our 'ought'-claims. So, to take a simple example, maybe you are wondering whether to take a bet on a horse in a race. You know that the horse has a great track record, and the odds are favorable, so given the information available to you, you ought to take the bet. In your context, it seems that you speak truly by saying that you ought to take the bet. But suppose that I know that the horse was recently severely injured, and that's a strong reason to decline the bet. It also seems that I can reasonably and truly say that you ought not to take the bet – or so the contextualist thinks, anyway. It's clear here why adding and taking away information from the modal base makes a difference to the truth-value of "you ought to take the bet": it affects which reasons are salient.

Rational requirements of the sort that we've been dealing with, however, are not concerned with responding to reasons. So it's less clear why adding or taking away information

from the modal base would make a difference to the truth-value of a claim about rationality. For example, take the sentence "it's irrational to simultaneously believe that it's raining and to believe that it's not raining." It doesn't matter how much background information you add or take away: it still seems true to say that having this pair of beliefs is irrational. The putative irrationality doesn't consist in failing to respond to some reason that might or might not be part of the information set; it is just inherent in the combination of attitudes. Moreover, 'ought' appears to have a kind of "super-objective" reading that depends on *all* the facts, whereas 'rational' seems at least in *some* important way limited to the subject's perspective. This might lead us to think that 'rational' is not as contextually flexible as 'ought' is – that it takes some particularly determined body of background information depending on the subject's perspective, whereas 'ought' can in principle take a range of other bodies of background information.

Originally, I thought that these observations showed that 'rational' – at least insofar as it's used in ordinary language to pick out the property that consists in conforming to coherence requirements – is not context-sensitive. But I have come to realize that this was too hasty.[221] What I now think the foregoing observations show is that the sentences that express the wide-scope requirements of rationality tend to be true irrespective of the semantic value of 'rational'. However, most of our talk about rationality is not explicitly wide-scope, and in the previous section I argued against interpreting it as such. So this does not show that there aren't different semantic values for 'rational', and that they don't make a difference to the truth of much of our talk. Indeed, I now think that there are, and that they do.

---

[221] Thanks to John Pittard and Jessie Munton for helping to convince me of this.

I will now explain how the contextualist semantics can be extended to talk about rationality to square all of this. The fundamental idea is that the ordering source for talk about rationality *is* satisfaction of the wide-scope requirements of rationality. In other words, the live worlds (as determined by the modal base) get ranked according to how well the agent satisfies the wide-scope requirements. So then, 'rationality requires S to Φ' is true iff in all the live worlds in which S satisfies the (salient) wide-scope requirements, S Φ's – that is, iff S's Φ-ing is necessary, given which worlds are live, for the requirements to be satisfied (to the greatest extent that they can be).[222] And 'rationality permits S to Φ' is true iff in some of the live worlds in which S satisfies the (salient) wide-scope requirements, S Φ's – that is, if S's Φ-ing is compatible, given which worlds are live, with the requirements being satisfied (to the greatest extent that they can be).

Let's apply this to an example. Let us return to the wide-scope formulation of the instrumental requirement:

(IR-Wide) Rationality requires of you that ((you intend end E ∧ you believe

that M is a necessary means to E) → you intend M).[223]

---

[222] This semantics forces us to give the English use of 'requires' what Broome (2013: 109-110) calls its "property" sense, i.e. that which states what is required to have a particular property, in this case that of being rational. This is to be contrasted with the "source" sense, in which we can talk of a particular source of requirements – rationality – requiring things of you, in the sense that it is rationality that *issues* these requirements. I think this is OK, since 'rationality requires' is something of a term of art, as Broome recognizes (Broome 2007: 361). Most of our ordinary talk about rationality does not use the term 'requires'. I will explain how to translate it into 'ought'-talk below.

[223] Of course, there's room to argue that this requirement should be more complicated, but I use a simple version of it for ease of explanation. The form of the explanation would go just the same way for a more complicated variant.

On our contextualist semantics, the semantic value of 'rationality requires' depends on context. However, the above sentence expressing the instrumental requirement itself will be tend to be true irrespective of the particular semantic value of 'rationality requires'.[224] This is because, no matter which worlds we restrict ourselves to, it will always be the case that in order to be in one of the worlds where you satisfy the rational requirements, you have to be in a world where you satisfy (IR-Wide) – since (IR-Wide) *is* one of the requirements of rationality that functions as the ordering source. So the sentence expressing the instrumental requirement will be true, for example, even if the modal base is completely empty.

What we now want to explain is how narrow-scope talk comes out true on our semantics as well. We will build up to explaining the truth of (1), which was, recall,

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

Let us begin with the following, rather unnatural, utterance:

(8) If you intend to start keeping bees (and you believe that purchasing a beehive is a necessary means to starting to keep bees), then rationality requires you to intend to purchase a beehive.

---

[224] An arguable exception is a context where it's in the modal base that you violate (IR-Wide). (Thanks to Bruno Whittle for pointing this out to me.) This raises significant complications that I cannot deal with here fully. I will just note that (i) in some situations, we may indeed want to talk about what would be the most rational thing for you to do *given* that you violate (IR-Wide); (ii) if we don't want to do this, we can exclude your violation of (IR-Wide) from the modal base, even if we know that you violate it (see section 3.4 above).

On our semantics, accepting (IR-Wide) actually *commits* one to the view that (8) expresses a truth. On the contextualist semantics, the antecedent of (8) restricts us to worlds in which you intend to start keeping bees, and believe that purchasing a beehive is a necessary means to keeping bees. *Once we restrict ourselves to those worlds,* the only worlds in which you satisfy (IR-Wide) – and thus the only worlds in which you satisfy the wide-scope requirements to the greatest extent possible – will be those in which you plan on purchasing (intend to purchase) a beehive. So, (8) comes out true.[225]

Now let us begin to make revisions. First, we can replace the "intending" language with more natural "planning" language, which I assume expresses the same thing in this context, so that it does not affect the truth-value of the utterance. So we get:

(9) If you plan to start keeping bees (and you believe that purchasing a beehive is a necessary means to starting to keep bees), then rationality requires you to plan to purchase a beehive.

The next step is to omit the parenthetical clause in the antecedent. Here we can rely on the maneuver already explained in the previous section. If the proposition that you believe that purchasing a beehive is a necessary means to starting to keep bees is already in the modal base – if it is something that we are assuming, holding fixed – then it does not need to be added to

---

[225] What if, given what we are holding fixed, intending the means would cause you to violate some other salient rational requirement? One can imagine multiple ways of going here. On one view, there will be some kind of fixed and general weighting of the different rational requirements for the purposes of the ordering source. In that case, there could be some special utterances of (8) which are false, in unusual contexts where given what is being held fixed, intending the means would cause you to commit some more gross form of irrationality. This raises the difficult question of how such a weighting would go. More plausible, in my view, is the claim that which rational requirements take priority in the ordering source is itself contextually determined by which rational requirements are salient. In particular, utterances like (8) make the instrumental principle salient since they explicitly bring out the connection between intended ends and intended means.

the modal base explicitly. The modal 'requires' will still be restricted to the worlds in which you have this belief. So as long as we're in such a context, we can omit it without affecting the truth-value of the utterance. We get:

(10)   If you plan to start keeping bees, then rationality requires you to plan to purchase a beehive.

The last thing we need is to be able to substitute 'you ought' for 'rationality requires you'. Since 'ought' has a parameter for an ordering source, we can conjecture that in the right context the ordering source for 'ought' (and not just for 'rationality requires') can be satisfaction of the rational requirements. In such contexts, 'ought' expresses the so-called "rational 'ought',"[226] as opposed to the *moral* ought, or the *legal* ought, or the *all-things-considered* ought, and so on. 'You ought to…' will then express the same thing as 'rationality requires you to…'. So we arrive at (1):

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

If the context is clear enough, we may be able to omit the parenthetical 'rationally'.[227]

---

[226] Dowell (2012: 283) makes this proposal, but does not develop it.

[227] I also think we should be able to omit the 'plan to' in the consequent, which also seems natural, if we are assuming that a failure to purchase a beehive would reveal a failure to intend to do so, and thus a violation of a rational requirement (cf. Broome 2013: 151). Again, we see here a way in which the contextualist semantics helps to mitigate purported problems with the wide-scope coherence account.

Someone might propose that we don't need a rational 'ought' to explain the truth of (1). Rather, we could take the 'ought' to be a so-called 'teleological' ought, where the ordering source is satisfaction of the agent's goals.[228] While I do not want to deny that there can be an 'ought' of this kind, such an 'ought' does not (*pace* the instrumentalist theory of rationality; cf. section 0.7 above) generalize for expressing requirements of rationality other than the instrumental requirement. For example, we want to be able to explain why it's OK to say:

(11)   If you believe that you ought to eat the cheese, then you ought to (intend to) eat the cheese

This can be explained by our semantics for the rational 'ought', given the presence of the enkratic requirement among the requirements that constitute its ordering source.[229] By contrast, it is not explained by a teleological use of 'ought'. The same point holds for expressions of instances of many other rational requirements.

We now have a theory on which the semantic value of 'rationality requires' (or 'rationally ought') is variable. As we have seen, statements of wide-scope requirements will tend to come out true irrespective of the semantic value of 'rationality requires'. By the same token, any sentence that describes some combination of attitudes that violate these wide-scope

---

[228] Compare Silk (2014: 6); Wedgwood (2006: 152).

[229] Of course, if we are contextualists about 'ought', we must say which semantic value of 'ought' features in the *antecedent* of the enkratic requirement: what kind of ought-belief one is required to line up with one's intentions. One initial proposal is that it is the super-objective ought: one is required to line up one's intentions with one's beliefs about what one objectively ought to do. However, so-called "Jackson cases" (Jackson 1991) call this into question. My tentative suggestion, indirectly inspired by Sepielli (2014), is that the relevant 'ought' for the enkratic requirement is something like the most subjective 'ought' about which you have a belief on the particular occasion. These issues deserve further exploration on another occasion.

requirements as rationally forbidden will come out true. What about statements of narrow-scope requirements? As we have seen, contextualism can explain the truth of an utterance like

(8) If you intend to start keeping bees (and you believe that purchasing a beehive is a necessary means to starting to keep bees), then rationality requires you to intend to purchase a beehive.

A sentence like (8) will tend to come out true irrespective of the initial context,[230] precisely because the antecedent (temporarily) adds propositions to the modal base that guarantee that the consequent comes out true. The same will be true of a more schematic version of (8) replacing the particular ends and means of this case with the generic schematic letters E and M, *as long* as we keep the ordinary English 'if…then'. However, by contrast, a narrow-scope statement of the instrumental requirement where the conditional is stipulatively the material conditional or some other logical operator will *not* come out true in virtue of the contextualist semantics, since the narrow-scope talk only comes out true in virtue of the conditional acting as a restrictor. (This contrasts with (IR-Wide), which as we saw above *does* come out true on the contextualist semantics even when the conditional is stipulated as material.) So although we have vindicated narrow-scope talk, there is an important sense in which we have not vindicated the narrow-scope requirement as traditionally understood.

Although (8) (as well as a more schematic version of it) will tend to come out true irrespective of the initial context, many other non-conditional (and thus trivially narrow-scope)

---

[230] Assuming that 'requires' is given its "property" sense – see fn. 222 above.

utterances about rationality may have different truth-values in different conversational contexts. Suppose I simply say

(12)   Rationality requires you to intend to purchase a beehive

If it is part of the modal base that you intend to start keeping bees, and that you believe that purchasing a beehive is a necessary means to this, then (12) is true. However, if these things are not part of the modal base for the utterance, then (12) is false – even if you in fact *do* intend to start keeping bees, and in fact *do* believe that purchasing a beehive is a necessary means to this.

One of the original motivations for the wide-scope view is the thought that when you (for example) intend an end, you may come to satisfy the instrumental requirement *either* by intending the means, *or* by dropping the end. Insofar as we have vindicated narrow-scope talk, have we abandoned that thought?

Consider again (8):

(8) If you intend to start keeping bees (and you believe that purchasing a beehive is a necessary means to starting to keep bees), then rationality requires you to intend to purchase a beehive.

In the case of (8), perhaps it's permissible to give up the intention to start keeping bees, rather than forming the intention to purchase a beehive. This may be true, but (8) doesn't say that it isn't. Having restricted ourselves to the worlds in which you *do* intend (at a time t) to start keeping bees, and in which you believe (at t) that purchasing a beehive is necessary for this,

it's then true that the only such worlds where you satisfy (IR-Wide) (at t) are those where you intend (at t) to purchase a beehive. It may be true that there are worlds where you permissibly give up your intention to start keeping bees (at t), but those are *ipso facto* just not worlds in which you intend (at t) to start keeping bees, and so the restricted claim doesn't say anything about them.

Indeed, notice that the contextualist semantics preserves the "symmetry" that is generally associated with wide-scope accounts. As noted just now, the semantics makes it true to say

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

But it will *also* make it true to say

(13)   If you don't plan to purchase a beehive, then you (rationally) ought not plan to start keeping bees

This too, given our semantics, is just a straightforward consequence of (IR-Wide). If we restrict ourselves to the worlds in which you don't plan to purchase a beehive, every remaining world in which you satisfy (IR-Wide) will be one in which you don't plan to start keeping bees. Given

the truth of both (1) and (13), it is clear that the account does not privilege forming the intention to take the means over dropping the intended end.[231]

3.7 Further puzzles resolved

a. Avoiding bootstrapping

The points just made help us to see how the present view avoids bootstrapping. On the old, non-contextualist narrow-scope view considered in section 3.2, an utterance like (12) above is true just in virtue of it being *true* that you intend to start keeping bees, and that you believe that purchasing a beehive is a necessary means to this. On the present view, by contrast, it is only on one reading of 'rationality requires' that (12) is true, even if those facts hold. Once again, we could use RR_R to mark the restricted semantic value of 'rationality requires' and RR_U to mark its unrestricted semantic value. We can only detach the claim that RR_R you to purchase a beehive; we cannot detach the claim that RR_U you to purchase a beehive. Moreover, the former claim is true regardless of what you intend, and the latter claim is false regardless of what you intend – the theory doesn't say that at the possible worlds that you have the intention, RR_U you to purchase the beehive. There is no bootstrapping here: the contextualist view does not in any sense allow that by intending to start keeping bees, you make it the case that you're rationally required to purchase a beehive.

Indeed, given our semantics, the truth of narrow-scope sentences like (8) just falls out of the truth of the wide-scope requirements. So they do not commit us to anything that the

---

[231] This bears back on the question explored in section 3.2 of whether the "symmetry" of wide-scope requirements is an advantage or a disadvantage. Let me register that (13) sounds pretty good to me! I think it is an advantage of an account that it can preserve its truth.

wide-scope requirements didn't *already* commit us to. This helps to explain at the most general level why it need not have the bad results that the original narrow-scope view did.

b.   *The synchronic/diachronic debate*

In my explanation in section 3.6 of how the semantics developed here vindicates narrow-scope utterances using the wide-scope semantics, I passed over an important point. Recall the explanation of

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

We add the proposition that you plan to start keeping bees to the modal base, and given this, the only way to satisfy the wide-scope instrumental principle is to plan to purchase a beehive. However, to be somewhat more explicit about the explanation, we might introduce 'at t' markers, as I did in section 3.6 above. You only violate the wide-scope instrumental principle, intuitively, if you plan to start keeping bees *at t* but do not plan to purchase a beehive *at t*. After all, the wide-scope principle is supposed to avoid forbidding you from giving up your end. So, if you plan to start keeping bees at t1, but do not intend to purchase a beehive at t2, then you may still be rational, if you have given up your plan to start keeping bees at t2.

This raises the vexed issue of whether rational requirements are synchronic – applying at a single time – or diachronic – applying over a course of time. It looks like I am assuming in the explanation given above that the instrumental requirement is synchronic. Made explicit:

(IR-Wide-Synchronic) Rationality requires of you that ((you intend, at t, end

E ∧ you believe, at t, that M is a necessary means to E) → you intend M at t).

However, one might think that in some sense the instrumental requirement is supposed to guide reasoning, and that reasoning involves forming attitudes on the basis of pre-existing ones. And one might think that a diachronic requirement captures this better. The problem, however, is that it's just not clear how to state a diachronic version of the wide-scope instrumental requirement. We could write:

(First Attempt) Rationality requires of you that ((you intend, at t1, end E ∧

you believe, at t1, that M is a necessary means to E) → you intend M at t2).

There are two massive problems with (First Attempt). First, it shares with a diachronic narrow-scope principle the result that giving up your intended end E at t2 is rationally impermissible. In this way, it betrays one of the main motivations for formulating the instrumental principle with wide scope.[232] Second, it is totally unclear what t2 is supposed to represent relative to t1. Picking any definite elapse of time between the two feels arbitrary.

Lord (2014a), a narrow-scoper, has a proposal for what a wide-scope, diachronic instrumental requirement could be. On his version (rewritten to match my schema), it says:

---

[232] I don't want to rule out the possibility that there might be *some* requirements of rationality which are to be understood as wide-scope, diachronic requirements of analogous form to (First Attempt). Perhaps this is the right way to understand rational requirements on belief updating like conditionalization, for example (cf. the initial thoughts I offer on this in the conclusion, part a). However, if there are such requirements, we will clearly have to accept that they *don't* have the sort of symmetry that wide-scopers have wanted to preserve for requirements like (IR): once you have taken the attitude in the antecedent (at t1), the only way to satisfy the requirement will be to form the attitude in the consequent (at t2): it is no longer an option to make the antecedent false, since t1 is now in the past.

(IR-Wide-Diachronic) Rationality requires of you that ((you intend, at t1,

end E ∧ you believe, at t1, that M is a necessary means to E) → (you intend M

at t2 ∨ you do not believe, at t2, that M is a necessary means to E ∨ you do

not intend, at t2, end E)).

This version of the diachronic wide-scope instrumental requirement restores the symmetry

associated with the wide-scope account. Lord proposes that we can just treat t2 as any time

that is t1 or later, thus also providing a determinate answer as to what it represents.

These are reasonable enough ways to make the diachronic principle palatable.

However, they are so reasonable that (IR-Wide-Diachronic) turns out to be logically equivalent

to (IR-Wide-Synchronic). Since Lord allows that t1 can be identical to t2, (IR-Wide-

Diachronic) entails (IR-Wide-Synchronic).[233] For the other direction, notice that it follows

from (IR-Wide-Synchronic) that you are required to make true the disjunction in the

consequent of (IR-Wide-Diachronic) at t2, *regardless* of what you believed at t1. So (given that

we are working with the material conditional), trivially you are required to (have them if

(material 'if') you had those previous attitudes at t1). So the diachronic version follows from

the synchronic one too. Similar points will hold for other conditional requirements.

Given this, I simply do not see how to state a diachronic wide-scope version of (IR)

that adds something new to the synchronic wide-scope version. Can we nevertheless

understand the wide-scope instrumental principle as guiding reasoning? In a way, I think the

fact that (IR-Wide-Diachronic) follows from (IR-Wide-Synchronic) shows that we can. You

---

[233] The more complex form of (IR-Wide-Diachronic) compared with (First Attempt) doesn't matter here,
provided one can substitute logical equivalents within the scope of 'rationality requires'.

are rationally required to satisfy (IR-Wide-Synchronic) not just now, but also at times in the future – since the synchronic requirement applies at every possible time. That means, I think, that you rationally ought to bring your mental states into line with (IR-Wide-Synchronic). To put the point in terms of our semantics, the ordering source for the rational 'ought' is not just whether you satisfy the synchronic requirements *now*, but also whether you satisfy them in the future.

With this in hand, we can revisit our semantics to deal with a potential worry about its explanation of the data. Recall that we wanted to explain ordinary utterances like

(1) If you plan to start keeping bees, then you (rationally) ought to plan to purchase a beehive

Since this sentence doesn't come with time-markers, it's not totally clear whether it means to assert something synchronic or diachronic. If it asserts something synchronic – that if you plan at t to start keeping bees, then you ought a t to plan to purchase a beehive – then it's clear how our semantics explains it. After all, once we restrict ourselves to the worlds in which you intend at t to start keeping bees (and have the means-ends belief we assume you have), the only worlds in which you are fulfil (IR-Wide-Synchronic) are those in which you intend at t to purchase a beehive. But what if it is supposed to assert something diachronic – that if you plan at t1 to start keeping bees, then you ought to plan at t2 to purchase a beehive? Then it seems like our semantics will not predict the truth of the utterance. For even restricting ourselves to worlds in which you intend to start keeping bees at t1 (and have the means-ends belief), it won't be the case that the *only* worlds in which you satisfy (IR-Wide-Synchronic) will be those at which you intend at t2 to purchase a beehive. You will also satisfy it in the worlds in which

you give up your intention to start keeping bees. So if (1) is meant to express something diachronic, we can't explain its truth.[234]

However, I think that if we make this kind of diachronic character explicit in (1), it ceases to seem so obviously true. Consider

(14)   If you plan just as of this moment to start keeping bees, then you ought, going forward and come what may, to plan to purchase a beehive

This makes the diachronic character explicit, but it ceases to seem obviously true. It seems subject to the basic objection that the wide-scoper was always so keen on: namely, might it not be OK to drop the plan to start keeping bees?

Does that mean that to be true, (1) must be explicitly synchronic in the sense of being limited to a single point in time? Not necessarily; for there are other things that our semantics, together with (IR-Wide-Synchronic) can explain as true. For example, consider

(15)   If you plan to start keeping bees, then you ought, going forward and so long as you stick to that plan, to plan to purchase a beehive

Our semantics can explain the truth of (15), since some future time at which you still intended to start keeping bees but did not plan to purchase a beehive would be a time at which you would violate (IR-Wide-Synchronic). Moreover, I don't think it's crazy to think that this could be roughly what's going on with the original utterance (1). (1)'s antecedent is that you plan to

---

[234] Thanks to Steve Darwall for raising this objection.

start keeping bees. It may be that what we add to the modal base when (1)'s antecedent is uttered is not just that you plan to start keeping bees *right now*, but that you will stick to this plan into the future. In other words, what we say is something like: given this plan of starting to keep bees and holding that plan fixed, you rationally ought to plan to purchase a beehive. There are a few different ways we could spell this out. We might say that the antecedent restricts us to the *time-indexed* worlds at which you plan to start keeping bees, and each of those worlds at a time is one where you will only be synchronically rational if you plan to purchase a beehive. Or we might say that since plans are generally future-directed and persist over time rather than being given up willy-nilly, the claim that you plan to start keeping bees causes us to add to the modal base the assumption that you have this plan not just right now but will continue to do so. All this illustrates that just because our ordering source requirements are synchronic does not mean that we can't make sense of talk about what you rationally ought to do going forward.

This, I think, is a pleasing result. We get it without having to endorse intuitively troublesome claims like (14), which again gives us a contrast with the traditional narrow-scope view. Earlier I said that our semantics renders true schematic versions of the narrow-scope requirements, as long as we keep the English 'if…then' rather than the material conditional. But in fact, we should be more precise. While it will vindicate synchronic versions of them, it will not vindicate *diachronic* versions of them. For example, it won't vindicate the claim

(16)   If you intend, at t1, end E, and you believe, at t1, that M is a necessary means to E, then rationally requires you, at t2, to intend M.[235]

---

[235] Adding (IR-Wide-Diachronic) to the ordering source obviously doesn't make our account vindicate (16), since, as proved earlier, it is just logically equivalent to (IR-Wide-Synchronic).

So once again, we see a sense in which the view comes apart from some traditional narrow-scope views.

c. Compliance and triviality worries

One objection to the wide-scope view of rational requirements focuses on its sometimes odd-sounding consequences about what counts as compliance with the requirement in question.[236] Take some specific instantiation of (IR-Wide), such as

(IR-Wide-Bees) Rationality requires of you that ((you intend to start keeping bees ∧ you believe that purchasing a beehive is a necessary means to starting to keep bees) → you intend to purchase a beehive).

(IR-Wide-Bees) requires you to satisfy the conditional embedded inside the brackets. Consequently, any instance of your not intending to start keeping bees, or of your not believing that purchasing a beehive is a necessary means to starting to keep bees, is an instance of your satisfying (IR-Wide-Bees). That may sound a bit odd. It suggests that for every intention or means-end belief you lack, you satisfy a corresponding requirement of rationality. By contrast,

---

[236] See Lord (2011, 2014a). In the latter Lord calls this problem the "real symmetry objection," but Lord's objection is really an entirely different objection to the one voiced by Schroeder and Kolodny considered in section 3.2 above. In fact, since Lord actually *agrees* that the correct requirements should permit either satisfying the consequent or ceasing to satisfy the antecedent, he is actually *precluded* from endorsing the Schroeder/Kolodny objection in principle.

a narrow-scope formulation can say that when you lack of one of these attitudes, the particular requirement of rationality does not apply to you.

I do not think that this objection should carry much weight. The satisfaction in question is vacuous, and all sorts of English talk that relies on vacuous satisfaction, or vacuous truth, can sound unnatural.[237] For example, suppose that no-one is going to the party, because it would be very boring. Then, in a vacuous sense, everyone who is going to the party is very excited about it. Likewise, in the same vacuous sense, everyone who is going to the party lives on Saturn, everyone who is going to the party eats Rafael Nadal for breakfast, and so on. But these ordinary sentences would strike ordinary speakers as downright bizarre in the circumstances. Needless to say, it's controversial whether the problem is semantic or pragmatic. At the very least, however, these sentences carry the implicature that someone is going to the party. Since, were that implicature true, the sentences would almost certainly be false, speakers will tend to judge them as false. Similarly, to say that you satisfy some specific instantiation of (IR-Wide) may often carry the implicature that you do so non-vacuously.

Moreover, one counts as satisfying the requirement when one lacks the antecedent attitude partly in virtue of the fact that the conditional that features in (IR-Wide) is material. For the purposes of stating the true wide-scope requirement, we just stipulate that the conditional we're using is material, regardless of what the ordinary English 'if…then' expresses. Nevertheless, most of our *talk* about what is conditionally required of you does not use the material conditional; as I've already argued, most of our ordinary *talk* is narrow-scope, but context-sensitive, with the ordinary-language conditional acting as a contextual restrictor as specified by the Kratzerian semantics. So, given that, there's another reason that it's not

---

[237] See also Wedgwood (2006: 149-50).

surprising that there is some unnaturalness in saying that one satisfies (IR-Wide) when one lacks the antecedent attitude.

Related considerations help to dispel the sorts of worries about triviality associated with the material conditional, which spelled part of the problem for wide-scope unqualified 'ought's (and thus the wide-scope *semantics*) in section 3.5. It might be worried that as long as we think that rational requirements are wide-scope, with the 'rationality requires' operator scoping over a material conditional, there will be lots of trivial rational requirements, which we can generate simply by putting a rationally impermissible attitude in the antecedent of the conditional. For example, we might be able to say that

(17)   Rationality requires of you that (you believe (it's raining $\land$ it's not raining) $\rightarrow$ you intend to eat a saucer of mud)[238]

Since rationality requires you not to believe contradictions, this appears to be vacuously true, as long as we read the conditional as material. Now, of course, the wide-scoper does have to admit that (17) is vacuously true. Still, the important point is that this does *not* transfer over to our ordinary *talk* about rationality. Given the contextualist semantics, the following utterance will not come out true:

(18)   If you believe that it's raining and it's not raining, then you (rationally) ought to plan to eat a saucer of mud

---

[238] Note that we've had to pick a different example from that which we used for the triviality problem for the unqualified wide-scope 'ought' in section 3.5. This is because there we could use anything which you (unqualifiedly) ought not do, whereas here we need something what it would be irrational to do (or believe).

For (18) to be true on our semantics, it'd have to be the case that, if we restrict ourselves to the worlds in which you believe it's raining and not raining, the most rationally ideal remaining worlds are all ones in which you intend to eat a saucer of mud. But that is just not true. So (18) comes out false. Therefore, the vacuous requirements don't have bad downstream consequences for our semantics, and nor will they play the same explanatory role that the non-trivial requirements do.

d. Why there is no "super-objective" reading of 'rational'

As noted above, 'rational' does not have the "super-objective" reading that 'ought' can have – the sense of 'ought' in which what one ought to do can be connected with facts completely outside one's grasp. As noted earlier, one might initially think that the explanation is that while 'ought' can take any body of information in the modal base, 'rational' can only take a body of information that is available to the subject. But this explanation seems wrong if 'rational' is context-sensitive in the way that I have proposed, since that explanation itself trades upon the possibility of adding particular propositions to the modal base so as to treat them as fixed. Such propositions can be added or taken away from the modal base according to the speaker's intentions, without any change in the information available to the subject.

However, the present account can offer a different explanation of why rational does not have a "super-objective" reading, albeit one that shows that this is a slightly misleading way of putting the point. On the present account, the difference between 'ought' and 'rational' is generated not by a difference in what can function as the modal base, but in what can function as the *ordering source*. Even if we do put all the facts in the modal base, that won't mean

that 'rationality requires' expresses the super-objective *ought*, requiring you to respond to all the reasons that are out there, because the ordering source for 'rational' (or the rational 'ought') is the set of wide-scope requirements, and those simply don't require you to respond to all the reasons that are out there; rather, they govern combinations of attitudes that you might have.

e. *Bolstering coherentism about rationality*

One important objection to a coherentist account of rationality is that a huge amount of our ordinary talk about rationality seems to presuppose that *individual* attitudes can be rational or irrational.[239] It does not seem to refer to combinations of attitudes.

The present semantics, however, shows that this objection is too quick. On this semantics, we can talk of rationality requiring individual attitudes of you, but only against a background where we have restricted ourselves to a limited set of worlds such that we are effectively holding other attitudes of yours as fixed (i.e., restricting ourselves to the worlds in which you have some particular attitude). Once we do that, we can just read off which individual attitudes are required to make you rational *given* the wide-scope requirements that you are under. If this is right, the coherentist need not give up on the truth of claims that say that a single attitude is rationally required of you. Along with the contextualist explanation of how rational requirements can be expressed with 'ought', this shows that the coherentist has a much greater capacity to explain ordinary talk about rationality than might otherwise be thought. So the present semantics helps to bolster the coherentist view considerably.

---

[239] See, e.g., Lord (2014b).

The view I have defended makes sense of narrow-scope talk. But it does so in a way that makes the wide-scope requirements explanatorily fundamental. It is the wide-scope requirements that give us the fundamental norms that constitute the ordering source.[240] This means that my account is still at heart a wide-scope account of the fundamental rational requirements, albeit one that preserves the truth of narrow-scope talk, and that avoids the wide-scope semantics.

3.8 Detachment rules

As I have stressed continually, wide-scope requirements are not in general requirements to have some attitude *simpliciter*. They are requirements which relate different possible attitudes, banning certain impermissible combinations of attitudes (or absences thereof). In this chapter, I have been explaining why it often nevertheless makes sense to talk of being rationally required to take some individual attitude, because we are holding other attitudes fixed in the conversational background. In a *linguistic* sense, this allows us to "detach": when one is rationally required to satisfy the conditional (one has attitude A→ one has attitude B), we can say that one is rationally required to have attitude B, provided that 'rationally required' has a semantic value that holds it fixed that one has attitude A.

---

[240] Relatedly, perhaps the wide-scope requirements are those which are true given a "source" sense of 'requires' – see fn. 222 above. If that's right, it might help to answer a worry put to me by Bruno Whittle: if the semantics for 'requires' appeals to an ordering source constituted by requirements which themselves have to be stated using the word 'requires', do we have some kind of problematic circularity or regress? Not if these statements can be understood as using 'requires' in the source sense. (One could also think of them as fundamental *demands*.) As I noted above, it is the property sense of 'requires' that is more clearly amenable to a Kratzerian treatment. So the semantics for 'requires' in the property sense utilizes requirements stated using 'requires' in the source sense as the ordering source, removing the potential circularity.

What if one rejects the notion of "source" requirements of rationality? Would Whittle's objection then spell trouble for the view defended here? Not necessarily. If one rejects the "source" conception, then one is likely to take the view that there is simply a property – rationality – and one can then say that the ordering source is simply the extent to which the agent exemplifies this property. Then the statements of requirements come out true, without our having to invoke the notion of a requirement in our semantics. (Thanks to Ralph Wedgwood for this suggestion.)

But in another sense, it still does not allow for the kind of detaching in the *logical* sense

that the simple, original narrow-scope view considered in section 3.2 did. It does not *follow*

from one's being required to make true the conditional (one has attitude A→ one has attitude

B), and one's actually having attitude A, that one is required (in any sense that is not implicitly

holding it fixed that one has attitude A) to have attitude B. That particular rule is called "factual

detachment", and it is universally agreed that it does not hold for wide-scope requirements (or

wide-scope 'ought's).

Still, there are some other, more plausible detachment rules that we should consider.

If such detachment rules hold, then this provides a further, and in one sense deeper, way in

which a wide-scoper can (and perhaps must) acknowledge that one may sometimes be

rationally required to take some individual attitude.[241] If detachment rules of some sort hold,

then the truth of this latter claim does not depend on 'rationally required' taking a semantic

value that implicitly holds some other attitude fixed in the conversational background. In other

words, the requirements to have the individual attitude would be unconditioned; they would

not be (implicitly) conditioned.

In particular, there is what is often called "normative detachment", which is licensed

by the so-called K axiom of standard deontic logic:

(K) $(O(p) \land O(p{\rightarrow}q)) \rightarrow O(q)$

Here 'O' stands for 'ought', but we can think of this axiom as also extending to requirements.

Unlike factual detachment, normative detachment does not allow us to derive a requirement

to have an attitude B from a requirement to make true the conditional (one has attitude A →
one has attitude B) plus the mere *fact* that one has attitude A. Rather, it allows us to derive a
requirement to have an attitude B from a requirement to make true the conditional plus a
requirement to have attitude A.

For my part, I find the K axiom and normative detachment extremely plausible at an
intuitive level.[242] The simple idea is that when you are required to make a material conditional
true, you can do that either by making the antecedent false or by making the consequent true.
But if you are required to make the antecedent true, then making the antecedent false is an
impermissible way of making the conditional true. So there is only one permissible way to do
what you are required to do, namely to make the consequent true. And when there is only one
permissible way to do something that is required of you, it's very natural to say that you are
required to do the required thing that way. So you are required to make the consequent true.

Broome (2013: 120-1, 127-8) denies the K axiom. Broome thinks that cases of the
following form are counterexamples to the K axiom. Suppose you ought to Φ, and you ought
[to make it true that] (you Φ → you Ψ). But suppose you do not Φ. Then, it seems possible
that you ought not to Ψ. So, to take Broome's concrete example, suppose that you ought to
exercise, and you ought (you exercise → you eat heartily). If you do not actually exercise, it
may not be the case that you ought to eat heartily.

I do not find this example persuasive. Specifically, I think that it can be easily defused
by contextualism in a way that preserves the K axiom.[243] There is a reading of 'you ought to
eat heartily' that is, as Broome claims, false. But this reading is one on which we hold it fixed

---

[242] Many writers take the K axiom for granted. Cf., e.g., Hare (1971) and Greenspan (1975).

[243] Wedgwood (2006) has some related discussion of how contextualism helps us to defend principles of
standard deontic logic.

in the background that you are not actually exercising. If we restrict ourselves to the worlds in which you don't exercise, then given the norms of prudence and healthiness which presumably constitute the ordering source on this occasion, it won't be the case that the top-ranked worlds are all ones in which you eat heartily. Hence, the claim 'you ought to eat heartily' will come out false, assuming we hold this piece of information fixed. But on the other hand, if we don't hold it fixed that you are not exercising (i.e., we do not restrict ourselves to worlds where you don't exercise), then the top-ranked worlds will be the ones in which you both exercise and eat heartily. Hence, 'you ought to eat heartily' will come out true in this context. And remember that either context can be operative, even if as a matter of fact you are not going to exercise.

This phenomenon of "norms of the second-best" should be familiar to contextualists. We can sometimes talk about what agents ought to do, *given* that they are not going to achieve the ideal – or even given that they are going to do something that radically deviates from what they ought to do (in a more unrestricted sense). This is what is going on in the example we discussed in 3.3 and 3.4:

(3) If you are going to punch your brother, then you ought to punch him softly

There is a norm of the first-best: you ought not to punch your brother at all. But there is also a norm of the second-best: given that you are going to punch your brother, you ought to punch him softly. Broome's example of exercise and hearty eating is just like this one. There is a norm of the first-best: in the unrestricted sense, you ought to eat heartily (as part of a plan where you also exercise). But there is also a norm of the second-best: given that you are not going to exercise, it's not the case that you ought to eat heartily (indeed, perhaps you ought not to eat heartily).

This preserves the K axiom by showing that the purported failure of the K axiom equivocates on a shift in the semantic value of 'ought'. The semantic value of 'ought' on which 'you ought to exercise' is true is the one where we are not treating it as fixed that you are not going to exercise. But on *that* semantic value of 'ought', 'you ought to eat heartily' is also true. So the example is not a counterexample to the K axiom, if one accepts the contextualist account.

Moreover, even if one rejects the contextualist account, I do not think that the counterexample is convincing. Broome's own preferred way of dealing with cases like (3) is to claim that they express wide-scope 'ought's. In these cases, he simply denies that it is the case that you ought to punch your brother softly. It is only the case that you ought (you are going to punch your brother → you punch your brother softly). But because *factual* detachment fails, we cannot get from this and the claim that you are going to punch your brother to the claim that you ought to punch your brother softly: you ought not punch your brother at all. But if that's the right treatment of *that* case, why not treat the exercise case similarly? The parallel treatment would be to claim that your mere decision not to exercise does not make it cease to be true that you ought to eat heartily. At most, what is true is some wide-scope claim, for example that you ought (you do not exercise → you do not eat heartily). But since factual detachment fails, one can deny that this entails that you ought not eat heartily, or that it's not the case that you ought to eat heartily. So it's unclear why Broome should endorse the claim that it's not the case that you ought to eat heartily, which is required for the case to be a counterexample to the K axiom. In this way, we actually see that the invalidity of factual detachment makes the K axiom, and thus normative detachment, *more* plausible, not less.[244]

---

[244] See Hare (1971: esp. 85-89) for a related and subtle discussion.

There are independent problems with Broome's decision to deny the K axiom. Let us suppose, not too fancifully, that the case is actually one in which, given that you are not going to exercise, you actually *ought not* eat heartily (and not merely that it's not the case that you ought to eat heartily). Nothing about this clarification of the case seems to affect our judgment that you ought to exercise, nor our judgment that you ought, if you exercise, to eat heartily. But if we accept both those judgments, and the judgment that you ought not eat heartily, then it seems that you are under a set of 'ought's that are not co-satisfiable. You cannot exercise, and satisfy the material conditional (you exercise → you eat heartily), and not eat heartily. So it seems that Broome is committed to saying that you are in a deontic dilemma, where 'ought's within the same domain conflict.[245] Yet Broome himself is suspicious of such dilemmas.[246]

Last, even if one did find Broome's counterexample convincing, we can weaken the K axiom so that it is not subject to these putative counterexamples:

(K*) $(O(p) \land O(p{\rightarrow}q) \land p) \rightarrow O(q)$

K* entails that, if you ought to Φ, and you ought (you Φ → you Ψ), and you do Φ, then you ought to Ψ. Broome's cases will not be counterexamples to this claim. K* still allows for a kind of detachment: when we have both the fact that you ought to satisfy the antecedent of a wide-scope ought and the fact that you *do* satisfy that antecedent, *then* we can derive the result

---

[245] Titelbaum (2015a) also notices this point, in a slightly different context.

[246] See e.g., Broome (2007). Of course, the natural way to respond to this is to say that this is not really a deontic dilemma, because the claim you ought to eat heartily is itself conditioned on your not exercising. But that just reinforces the inadequacy of Broome's treatment, on which all three 'ought's hold on the same reading of 'ought' (which recall, is what is needed for this to be a genuine counterexample to the K axiom), to the agent at the same point in time. These three 'ought's are not co-satisfiable, and to be in a situation where you are under two or more 'ought's that are not co-satisfiable just is to be in a deontic dilemma.

that you ought to satisfy the consequent. I conclude that normative detachment remains very plausible even in light of Broome's counterexamples, and that even if one rejects it, some similar weaker principle is plausible.

We can now return to our discussion of rational requirements. Given wide-scope rational requirements and the truth of the K axiom (i.e., the validity of deontic detachment), will we be able to generate lots of unconditional requirements to have individual attitudes? One might be skeptical about whether we will. After all, the K axiom generates a requirement to have some individual attitude only when we *begin* with a requirement to have some other individual attitude (plus a wide-scope requirement relating the two attitudes). So it does not seem to generate requirements to have individual attitudes solely out of wide-scope requirements. However, Titelbaum (2015a) shows that, under particular circumstances, one can use the K axiom to derive requirements to have individual attitudes from *two* wide-scope requirements.[247] It's worth being clear that not just any two wide-scope requirements will do; they have to be of a particular form for Titelbaum's demonstration to go through.[248] So the instances in which a wide-scoper is committed to unconditional rational requirements to have an individual attitude may be somewhat limited. Still, I am happy to acknowledge that such cases may be possible.

---

[247] See also Brunero (2010: 40-42).

[248] Specifically, one of the requirements has to be a form whereby the deontic operator that states the requirement (taking wide-scope over the material conditional) has to appear *again* as part of the content of the belief that appears in the antecedent of the requirement. Titelbaum's example is a version of the enkratic principle: Rationality requires of you that (You believe that rationality requires you to Φ → You Φ). But it's controversial whether this is the right statement of the enkratic principle. Broome's canonical statement of the enkratic principle uses 'Rationality requires' as the operator outside the scope of the conditional, but 'ought' as the operator within it. His claim is that you are rationally required to do what you believe you ought to do, not that you are rationally required to do what you believe you are rationally required to do (indeed, Broome actually denies the latter, Broome p.c.). And it's crucial to Titelbaum's demonstration that the requirement be of the second variant.

A different, and perhaps, more widespread, source of unconditional rational requirements to have particular attitudes may come from "necessary detachment". Necessary detachment allows us to derive a requirement that you have attitude B from a requirement that (you $\Phi \rightarrow$ you Ψ) and it's being *necessary* that you Φ. Now, it might initially seem hard to think of cases where it is necessary that you Φ. But some may propose that we should treat the past as effectively necessary for the purposes of deontic logic, since it is now unchangeable.[249] If that's right, and there are diachronic requirements of rationality, then it looks like we can derive unconditional requirements to have particular attitudes. Suppose there is a diachronic requirement that (you have attitude A at t1 $\rightarrow$ you have attitude B at t2). If t1 is now past, then the only way to satisfy this requirement is by having attitude B at t2. So one might think that one is now simply required to have attitude B at t2.

In section 3.7, part b, I argued that the kinds of requirements of rationality we have been considering so far as not best understood as diachronic. Though there are diachronic requirements of rationality that trivially follow from their synchronic variants, there diachronic requirements themselves have wide-scoped *consequents*. So they do not allow us to derive requirements to have individual attitudes *simpliciter*. That said, there may be some other diachronic requirements of rationality.[250] So, if necessary detachment is valid and the past should be treated as necessary, then there is another way in which we may be able to derive unconditioned rational requirements to have particular attitudes.

---

[249] Cf., e.g., Greenspan (1975: esp. 265). Thanks to Ralph Wedgwood for making this point to me.

[250] See the conclusion, parts a and b.

Chapter 4
Believing at Will as a Failure of Coherence

4.1 Introduction

In this chapter, I will be trying to show that a focus on the coherence requirements associated with belief – the sort that I have defended in the dissertation so far – can help to resolve a classic philosophical puzzle. So we will see a case study of how such requirements shed light on an important extant debate in the literature. The puzzle is this: why is it so hard – arguably impossible – to believe something directly and consciously on the basis of pragmatic considerations? That is, why is it so hard to "believe at will"?

I'll say what some of the important terms just mentioned mean in a minute. Before doing that, it'll help to have a case on the table. Suppose I offer you $1,000,000 to believe that the text of this sentence is green in color. Presumably you would like $1,000,000, and presumably you care about having $1,000,000 *much* more than you care about having a true belief about the color of the sentence that you just read. But try as you might, you will have *severe* difficulty believing that the sentence is green in color.

The consideration that believing the text is green will get you $1,000,000 is a paradigm instance of a pragmatic consideration. We mustn't get hung up on the conventional association between "pragmatism" and the pursuit of self-interest, though. To philosophers – lovers of truth – it may seem that there is something disreputable about sacrificing the pursuit of truth for one's own material wealth. But we can easily change the example so that, say, believing the text is green will win $1,000,000 for a charity that will use it to save hundreds of lives. Surely believing a minor, unimportant falsehood for the sake of saving hundreds of lives is a perfectly honorable sacrifice. Yet this modification of the example doesn't make it any easier to believe the proposition in question. There are even cases where the benefits brought about by the

belief are themselves epistemic.[251] Perhaps an angel (or a demon, depending on your perspective) makes you the offer that if you believe that the text is green, the rest of your beliefs from now on will all be true (or even that they will all count as knowledge, if you want that too). Still difficult to believe.

In what follows, then, when I talk about "pragmatic" considerations, I mean to use term broadly, such that the extension of "pragmatic" is broader than "self-interested", though cases where the considerations are self-interested probably provide the most sociologically interesting and widespread examples. What makes a consideration pragmatic in this wide sense is roughly just that it is not evidential, in the sense that it is not connected with evidence for believing the proposition in question. That's only a rough characterization because there are some considerations that are *putatively* evidential, whereby the subject *takes* them to be evidence for their beliefs, but are in fact not (good) evidence. (This apparently mundane fact will turn out to be important later.) I wouldn't want to call those considerations pragmatic. So pragmatic considerations are, slightly less roughly, those which are not even putatively evidential. I take no stand on whether pragmatic considerations amount to *reasons* for belief, properly understood.[252] I am deliberately using the slightly vague term 'consideration', rather than 'reason', in order to stay neutral on this.

---

[251] See Foley (1993: 21); Railton (1994: 73-4); Harman (1999: 101).

[252] Shah (2006) and Parfit (2011: Appendix A), amongst others, contend that pragmatic considerations cannot constitute reasons for believing at all; at most, they are reasons to do something that issues in belief. For a defense of genuine pragmatic reasons for belief against these sorts of arguments, see Reisner (2009). In this chapter, I am not directly focused on whether pragmatic considerations deserve the honorific 'reasons for belief' – instead, I will ultimately be interested in which ways of responding to pragmatic considerations involve one in (ir)rational incoherence. These issues may ultimately be linked in some way, and I hope to be able to pursue that question elsewhere.

Perhaps the best characterization is that pragmatic considerations in our sense are those connected with the value of being in the state of belief (they are "state-given"), rather than considerations to do with the "belief-worthy" features of the proposition that is the object of belief (those are "object-given").[253] But even that will be tendentious.[254] I don't want to try to settle all of this here, so instead rely on your by now having a good sense of what I mean by "pragmatic considerations" as, roughly, considerations that are not even putatively evidential with respect to the proposition in question.

I have three, highly interrelated, contentions in this chapter. First, I argue that philosophers have tended to overstate the generality of the claim that believing at will is difficult (or, according to some, impossible). Even when believing at will is distinguished from its nearby neighbors, there are certain cases which are plausibly regarded as genuine (albeit in some ways nonparadigmatic) instances of believing at will. Second, I aim to cast doubt on the popular thought that the difficulty of believing at will can be explained by the substantive norms associated with belief – that belief is subject to a norm of truth, that it "aims" at truth, or even that only evidential reasons can be normative reasons for belief. I argue that even if these claims are true (a point on which I want to stay noncommittal), they are at best incomplete as an explanation of why believing at will is difficult, and fail to adequately explain the difference between the paradigm cases where believing at will is very hard and the other cases where it is less so. Third, I suggest that we can do better by examining the coherence requirements associated with belief. I argue that believing at will, in the paradigm cases, would require one to be incoherent in a way that is transparent to oneself – and, picking up on a

---

[253] Cf., e.g., Parfit (2011: Appendix A); Olson (2004); Piller (2006).

[254] See Schroeder (2012) for worries about equating the distinction at issue with the object-given/state-given distinction.

more general feature of coherence requirements identified in section 2.1 – that this kind of *transparent* incoherence is hard for agents to sustain. This strategy provides both a fuller explanation of the difficulty of believing at will in the paradigm cases, and a better explanation of the difference between these cases and the overlooked cases in which believing at will is easier.

The case study of believing at will illustrates the theoretical fruitfulness of distinguishing substantive normative reasons from coherence requirements, as I have been doing throughout this dissertation, and how the latter (and a notion of rationality that gives pride of place to them) are poised to play *explanatory* roles that the former are not – a point that I also raised in section 2.1 and now want to illustrate.

4.2 What is it that's difficult?

I said that it's difficult to believe something *directly and consciously on the basis of* pragmatic considerations. What does that mean? Again, it will help to characterize this description by contrast with other (sometimes nearby) cases in which it *is* possible to respond in some way to pragmatic considerations in belief-formation.

(a) Wishful thinking

Start with cases of the very common phenomenon of "wishful thinking". As Thomas Kelly has astutely observed,[255] many cases of wishful thinking are not ones in which pragmatic advantages of *believing* p play the crucial role. Rather, the paradigm case of wishful thinking is

---

[255] Kelly (2002: 177); see also Williams (1973: 150). For a parallel contrast in a different setting see Worsnip (forthcoming-b).

one in which I believe p because I desire or wish that p *actually* be true. This is not necessary an instance where *believing* p is pragmatically beneficial, as the following example should bring out. Suppose that I am embarking on a new business venture with my business partner Dave. I have put a lot of my own money into the joint venture already, and Dave has been delaying putting in his share. Since I've put a lot of my own money into the venture, and this money will be wasted unless Dave puts in his share, I strongly desire that Dave come through with the money. As a result, I may wishfully believe that Dave is going to come through with the money – even if I have extensive evidence that Dave is a thoroughly unreliable individual, and that his delay-tactics are a classic sign that he will never come through with the money. But my having this wishful belief might in reality be very harmful to me. Perhaps since I keep on falsely believing that Dave is going to come through, I continue to pour my own money into the project, wasting more and more money in the process. So this is not a case where my having the belief is a result of the *belief itself* being beneficial to me. It is a case where I have the belief because I want the content of the belief to actually obtain.

Still, I think there are other cases in which wishful thinking can be an indirect response to the pragmatic advantages of the belief itself. Suppose that you're a philosophy graduate student looking for a dissertation topic, and your advisors have stressed to you that it's important for getting a job that you defend an original view. And suppose also that you realize that no-one has yet defended the view that (say) Kant was an expressivist about normative language. It's not too much of a stretch of imagine that, without having any kind of explicit thought of the form "I'd better defend that view, because it's not taken yet, and I need something original", you might find yourself defending, and *genuinely believing*, this view, in a way that is subconsciously influenced by your recognition of the fact that believing this heretofore-undefended view would help you to get a job. (Of course, you could just pretend

to believe the view while secretly disbelieving it; but you might well be less effective if you tried this.) In this case, it is not that you (antecedently) want it to be true that Kant was an expressivist: you might not really care one way or the other. Rather, what subconsciously produces the belief is the fact that *believing* that Kant was an expressivist is pragmatically beneficial to you, in helping you to secure a job. This still seems like wishful thinking in a good sense of the term, but here the causal influence on your belief is at least primarily down to a desire to secure the benefits of believing, rather than a desire that the propositional content of the belief obtain.

What is clear across both kinds of case of wishful thinking is that when one succumbs to wishful thinking, the influence of these pragmatic considerations is subconscious. Often, when one is thinking wishfully, one will think that one has good evidence for one's belief, relying on placing great weight on the evidence that supports one's belief and dismissing or downplaying the evidence against it. Moreover, the process of arriving at a wishful belief is typically subconscious from start to finish: it is not the result of *conscious* manipulation of one's own doxastic states.

(b) Plotting against yourself

Reflection on cases of wishful thinking (and their ubiquity) might lead one to propose that what is always hard is *deliberately* responding to pragmatic reasons in belief-formation. But this is still not right. There are cases – let's call them, following Foley, cases of "plotting against yourself"[256] – where you can deliberately and consciously set off a process that will indirectly

---

[256] Foley (1993: 17). See also Williams (1973: 149-51); Shah (2006: 495-6).

bring you to form the belief. In setting off the process initially, you can be completely consciously aware of what you are doing. For example, consider a variant of the case at the start where I offer you $1,000,000 to believe the text is green. Suppose you have the opportunity to get (easily reversible) laser eye surgery that makes black text appear green, to undergo hypnosis to get rid of the memory of undergoing the surgery, and to ensure in advance that you will look at the sentence in question before you've had the opportunity to figure out that all black text now appears green to you (all for far less than the $1,000,000 you'll gain). Knowing this, you may execute the plan, with the aim of getting the $1,000,000 initially in mind, and so come to believe that the text is green.[257] In so doing, you again manage to respond to pragmatic considerations.

Perhaps now it might be suggested that what makes the case possible is that at the time that you form the belief, you no longer realize that you began the process by being motivated by pragmatic reasons.[258] Again, however, this isn't right. Suppose that your method for getting yourself to believe that the text is green is to join the Green Text Conspiracy Hawkers' Association (GTCHA), which propagates the theory that all purportedly black text is green, and that we only perceive it as black because we are trapped inside an evil government-sponsored machine that dulls our senses. Ahead of joining the association, you have very little sympathy for this view, but you do know that you're easily persuadable, especially by charismatic orators, and you've heard that GTCHA has several of those. Now let's say that once you've joined, you are indeed persuaded that the text is really green. You may still remember that originally you took there to be no evidence for the claim that the text

---

[257] As Bovens (1995) draws our attention to, there are ways of indirectly manipulating your beliefs that don't involve anything this extreme – in particular, beginning to act as if you believe the proposition that you want to believe.

[258] Cf. Foley (1993: 17-18), who says that plotting yourself requires you to cover your tracks.

is green, and that that your only reason for joining was to get the $1,000,000. Still, you now believe that the text is really green – you look back on your past self, the one who judged that there was no evidence for this claim – and think, *what a fool I was to think that! Thank God that strange reward got me to join the organization that has now taught me the real truth!* (Maybe your past self even *knew* that this was what your future self would think of your past self, and nevertheless happily set the process in motion.)

So you can even recognize the genetic claim that your belief was causally produced as a result of your responding to pragmatic reasons.[259] What seems hard, still, is to be consciously basing your belief on pragmatic considerations *at the time that you form the belief*: to sincerely say, in a clear-eyed way, "I have decided to believe p, and my grounds are that believing p would be pragmatically beneficial." This is what I was trying to get at when I said that we have difficulty believing things directly *and* consciously on the basis of pragmatic reasons. Some instances of wishful thinking might be (at least in an important sense) direct: the desire to believe p simply produces belief in p, with no causal intermediary.[260] And some instances of plotting against yourself are executed in a calculated and (at least initially) conscious manner.[261] What seems hard is believing something directly *and* consciously on the basis of pragmatic considerations, such that one is conscious *at the time of belief-formation* that one's grounds for believing are pragmatic.

---

[259] Compare Williams (1973: 148); Adler (2002b: 58), who both seem to assume that this isn't possible.

[260] See Hieronymi (2006: 48).

[261] As Hieronymi (*ibid.*: 48-49) again points out, it can't be that anything that requires a multi-step process thereby counts as involuntary for that reason alone: "surely whatever sense of "voluntary" divides believing from raising one's right hand also divides believing from preparing dinner."

Many philosophers describe the problem as one of "believing at will". I am willing to accede to that usage and stipulatively use "believing at will" to refer to believing something directly and consciously on the basis of pragmatic considerations. But my topic here is not really what makes for genuine *freedom* or *agency* in belief, or what the preconditions of *responsibility* for belief are. Those are interesting questions,[262] but I already have enough on my hands. When I talk about "believing at will," I just mean believing something directly and consciously on the basis of pragmatic considerations. I hope that stipulation is justified by my argument so far that this is the phenomenon that is distinctively difficult for us, in contrast to cases of wishful thinking and plotting against oneself.

4.3 Is believing at will impossible? Some possible counterexamples

I have said that believing at will (in the stipulative sense just specified) is hard. But are there some cases in which it is possible?

Philosophers underestimate the extent to which ordinary folk often talk as if they are capable of a certain amount of believing at will. The exact phrase "I choose to believe" gets 393,000 hits on Google. Of course, we can say that all these usages either misuse "choose" or misuse "believe", but I think we should at least be open-minded enough to countenance the possibility that some of these speakers are not misusing terms. Here are some examples – not supposed to be exhaustive – that I think are at least worth taking seriously:

1. *Some cases of religious belief.* Some religious believers talk about their belief in God as a deliberate choice based on (e.g.) a need for solace and comfort in their lives. Likewise,

---

[262] For treatment of these issues see amongst many others Owens (2000), Shah & Velleman (2005) and Hieronymi (2006).

religious proselytizers often appeal to these considerations when trying to convert people.

2. *Trusting the testimony of friends.* Some people talk about their believing the testimony of friends as a decision, based on their duty of loyalty.[263] For example, suppose that a close friend is accused of a crime. I might say that I have to believe he is innocent, because to do otherwise would be disloyal.

3. *Believing in oneself.* Often believing that one is going to succeed at something is presented as a decision. The classic example is a sports team, where the coach says "you have to believe that you are capable of doing this – otherwise you'll never do it." This seems to be a case where a pragmatic consideration is presented in favor of believing, and it doesn't seem a stretch to me to think that the team members can sometimes consciously decide to believe for this reason. Such cases have the unusual and interesting feature that once one does decide to believe p, p does actually become more likely.[264] So one's decision to believe p for pragmatic reasons does actually create some non-pragmatic evidential reasons to believe p. But we shouldn't overstate this. The evidential reasons created by the decision to believe p may be quite weak. Suppose we are a basketball team down by 5 points with 7 seconds left on the clock. That kind of comeback has been made before, but is on balance very unlikely, given our evidence.

---

[263] Some philosophers have argued for the normative claim that friendship imposes obligations to trust one's friend (even when that involves deviating from the evidence). See Meiland (1980), Keller (2004) and especially Stroud (2006).

[264] So some seem to treat this as a special case where believing at will is possible, for this reason. See James (1896); Velleman (1989: ch. 5); Harman (1999: 16); Millgram (1997: 34).

It's certainly more likely conditional on our believing we're going to win than it is conditional on our not believing we're going to win. So (since we want to win) it is worth trying to believe that we're going to win if we can. But it's still on balance pretty unlikely that we'll win even conditional on our believing that we will. Though our believing we're going to win gives us some evidential reason to believe we'll win, the reason is definitely not strong enough to make it the case that on balance, our evidence supports believing that we'll win.

4. *Deciding whether to terminate inquiry (in belief) or continue inquiry (in a state of suspended judgment).* Sometimes, when one has been inquiring into whether p, one can reach a point where one faces a decision about whether to now go ahead and believe p, or whether to keep on inquiring about whether p. Such cases are subtle. In these cases, the considerations one has adduced in favor of believing p are themselves evidential. However, the decision one faces in such cases is not necessarily, phenomenologically, one of thinking "do the evidential considerations I have adduced so far meet some pre-determined bar that is sufficient for believing?" Rather, it can be, in some cases, more like thinking, "where shall I set the bar of how much evidence I require in order to be willing to believe?" And this decision about where to set the bar can be determined by pragmatic considerations.[265] One can think, "it's going to be very costly and difficult to continue inquiry about this matter, and I'm getting tired, and ultimately if I make a mistake it's not going to be the end of the world. So I'm just going to stop questioning p and believe it." Or one can think, "it's easy to continue inquiring about this matter, and it's

---

[265] Indeed, once we focus on the question of where to set the most *basic* standards about how much evidence to require, it's not obvious that such a question even *could* be settled by evidential considerations themselves.

really important that I get this right. I'm not ready to believe p yet." This feels (to me at least) like an immediate decision about whether to believe p, on which one could go either way, depending on one's appraisal of the pragmatic considerations bearing on future inquiry.[266]

Of course, one may be drawn to immediately start adducing reasons why these cases are different from the paradigm cases where believing at will seems overwhelmingly difficult (like the case where I offer you $1,000,000 to believe the text is green). I don't deny that there are differences, and it is certainly a desideratum of a good explanation of the data that it identify some such differences. We will want to explain why these cases seem somehow unparadigmatic later. Ultimately, however, these examples nevertheless seem at least like good candidates for being cases in which subjects are able to believe something directly and consciously on the basis of pragmatic considerations – i.e., in our stipulative sense, to believe it at will.[267] Perhaps if we cannot arrive at any explanation as to why these cases could be possible while the paradigm cases are not, we will have to write them off as not counting as genuine cases of *belief*, as opposed to some other attitude. But I think that is to be avoided if possible. In the friend case (at least suitably developed), for example, it's important to us that we say "I must believe my friend", rather than, say, "I must act as if I believe my friend". So I think it is at least worth seeing whether there is an account which will allow these examples to count as beliefs while still explaining why the attitude seems easier to form at will than in the paradigm cases. I will eventually claim to provide such an account.

[266] See Shah & Velleman (2005: 511-514, 517); Schroeder (2012: esp. 470-472).

[267] Compare Adler (2002b: 10), who urges us to ignore these cases, focus on the paradigmatic ones, and then assume that the best theory will extend the right theory for the paradigm cases to the nonparadigmatic ones.

Regardless of that, however, there's no doubting that in some paradigm cases – like the one we began with where I offer you $1,000,000 to believe that the text is green – believing at will is at best exceptionally hard. So there is certainly *something* to explain.

4.4 *An explanation in terms of substantive norms?*

One particular kind of proposal has been particularly prominent in trying to account for the difficulty of believing at will. The proposal is that this difficulty has something to do with the fact that belief aims at truth.[268]

What does it mean to say that belief aims at truth? The metaphor is notoriously hard to unpack. One bad proposal, at least in the current context, is that beliefs actually do track the truth.[269] Since many beliefs are false, this claim could only be that beliefs generally *tend* to be true.[270] But this claim – whether it is true or not, and it's a bit hard to assess such a huge empirical generalization – does not seem like it could on its own explain why believing at will is so difficult. Even if beliefs tend to be true, many beliefs are in fact false, and in many cases of false belief there is no puzzle about how the false belief is possible. The present proposal does nothing to show why believing at will in particular should be difficult. Moreover, it is just as hard to believe a true proposition on the basis of pragmatic considerations as it is to believe a false proposition on the basis of pragmatic considerations. Holding everything else constant, the actual truth-value of the proposition doesn't seem to make a difference to the difficulty at all.

---

[268] See Williams (1973), Foley (1993: 16), Adler (2002b), Hieronymi (2006), Velleman (2000).

[269] Cf. Williams (1973: 148): "If I could acquire a belief at will, I could acquire it whether it was true or not."

[270] Cf. Velleman (2000: 254).

A better, and more serious, proposal, is that what it means to say that belief aims at truth is to say that belief that it is subject to a *norm* of truth, the norm that one ought to believe p if and only if it is true. There are obviously some semantic values of 'ought' on which that claim comes out false (there is clearly some sense in which it's not the case that you ought to believe truths for which you have absolutely no evidence, for example). But nevertheless, there is a "super-objective" value of 'ought', perhaps, on which it comes out true (see section 3.7 part d above). If you don't like that idea about 'ought', you can substitute the notion of correctness: a belief counts as correct iff it is true. As a number of philosophers have plausibly contended, this notion of correctness is normative.[271]

How is this proposal supposed explain the difficulty of believing at will? The idea here cannot be that violating this norm is in itself difficult, for it clearly isn't: again, false beliefs are often easy. So consider yet another proposal: believing at will somehow amounts to flouting this norm *by one's own lights*. This gets closer to the truth, I think, but is still wrong as it stands. When one comes to believe some proposition, one believes that proposition to be true. If I could believe p at will, that would involve coming to regard p as true (at will). It would not involve coming to believe p while still regarding p as false. In fact, I need not have even previously regarded p as false. I may have been previously non-committal with respect to p. So believing at will is not in any clear sense putting myself into a state where by my own lights I flout the norm of believing what is true.

Perhaps we can do better in making good on what seems right here by starting to think about *evidence*, and its connection to truth. Believing at will is coming to believe something directly and consciously on the basis of pragmatic considerations. By definition, pragmatic

---

[271] See, e.g., Wedgwood (2002b); Shah (2003); Shah & Velleman (2005); Gibbard (2005).

considerations are non-evidential. Evidential considerations have to do with *markers* or *indicators* of truth: so, an evidential reason to believe p is a consideration that is a reason to believe p because it indicates to some degree that p is (in some sense) likely to be true.[272] That means that to believe p on the basis of pragmatic considerations is, by definition, to believe something on the basis of something that does not indicate that p is true. So to do so directly and consciously amounts to something like conscious indifference to whether one's belief is true. And that is in some way difficult.

The phrase "something like conscious indifference to whether one's belief is true" here is vague for a reason. For the word "indifference" is too strong. Allowing one's beliefs to be influenced by a truth-irrelevant factor does not really amount to *indifference*, properly speaking, to whether one's beliefs are true. One might prefer that one's beliefs be true, while still also caring about their pragmatic effects. But it seems at least roughly correct to say that to be influenced by pragmatic considerations is to consciously compromise one's effort to conform to the truth norm, that is, the norm of believing a proposition iff it is true. Following one's evidential reasons is one's best attempt to conform to the truth norm, and allowing anything else into the picture can only interfere with that attempt.[273] So by believing consciously and directly one the basis of pragmatic reasons, one is consciously compromising on the goal of conforming to the truth norm.

There's a sense, for me at least, that there's at least something right about this idea. But as it stands, the explanation of why believing at will is so difficult is still radically incomplete. There are several (groups of) questions that it leaves unanswered.

---

[272] See, e.g., Harman (1999: 17); Foley (1993: 17).

[273] Cf. Wedgwood (2002b: 276).

First, *why* should it be difficult to consciously compromise on the goal of conforming to the truth norm? This surely can't follow simply from the fact that there *is* a norm of believing a proposition iff it is true. There are many norms that one can consciously compromise on. Take your favorite theory of normative ethics: you (morally) ought to Φ iff Φ-ing has property F. It surely doesn't follow from that being a true norm that one can never consciously compromise the goal of following the norm. For example, if it were true that you ought to Φ iff Φ-ing maximizes general utility, it wouldn't thereby follow that one can never act so as to consciously compromise the goal of maximizing general utility.

It may be replied that the truth norm is *constitutive* of belief. That is, it isn't *just* the case that a belief is correct iff it is true; rather, this fact is part of the very essence of what it is for something to be a belief: part of what it is for something to be a belief is for it to be subject to the truth norm.[274] I don't think this really demystifies things. What is it about a norm being constitutive that makes it impossible to consciously compromise on the goal of fulfilling it? In fact, it seems that there are constitutive norms that it is *not* difficult to consciously compromise on. First example: promising. It's plausible that it's constitutive of the act of promising that such promises are subject to a norm whereby it is wrong to break them without excuse. But it is easy to consciously compromise on fulfilling this norm. Second example: assertion. Many think that it is constitutive of the act of asserting that it is governed by a norm (depending on who you ask, a norm of truth, or a norm of knowledge). But again, it's still easy to lie.

Similarly, someone might suggest that the truth norm is one that agents are all implicitly committed to in virtue of their status as believers.[275] So, in believing p, one commits

---

[274] Adler (2002b); Shah (2003).

[275] Cf. Hieronymi (2006, 2009); Shah (2003).

oneself to regulating one's belief in accordance with the truth norm. Perhaps I must *grasp* or *accept* the truth norm to count as (reflectively?) believing.[276] But even if I am so committed or in such a state of grasping or acceptance, why does it follow that I can't consciously fail to live up to my commitment, or to the norm that I grasp or accept? Again, the analogy of promises looms. Promising, in a very clear sense, commits one to keeping one's promise unless one has excuse. But I can consciously fail to live up to this commitment. Likewise with other norms that I grasp or accept.

I am not saying that these questions could not possibly have answers. I am just saying that, until such answers are provided, the explanation of why it is so difficult to believe at will is incomplete.

Second, why is it possible to respond to pragmatic reasons indirectly in the ways noted in section 1? Embarking on a plot against yourself still amounts to consciously compromising your future self's satisfaction of the truth norm. Why is this kind of compromise possible, but compromise on your present self's satisfaction of the truth norm impossible?

Third, why is it at least arguably possible to believe at will in particular cases – the ones that I drew attention to in section 4.3? Recall that there were four (types of) examples: religious belief, trust in friends, believing in oneself, and deciding whether to terminate or continue inquiry. The first three examples all seem to me to involve conscious compromise on the goal of believing the truth. That isn't to say that religious beliefs are all false, that trust in friends is always misplaced, and that belief in oneself is always mistaken. But even in the paradigm cases where it is hard to believe at will, as we have seen, the problem is not that the belief in question has to be false. Rather, the problem is that to the extent that one lets oneself be influenced by

---

[276] See esp. Shah (2003: 470).

non-evidential factors, one is allowing for interference with one's *best attempt* to conform to the truth norm (by being guided by the evidence only). This seems to be the case in the first three examples just as much as it is in the paradigm cases.

On the other hand, the final example – that of deciding whether to terminate or continue inquiry – does not obviously involve a compromise on the goal of conforming to the truth norm. But that is only because the goal of conforming to the truth norm, as William James pointed out long ago,[277] actually factors into two goals, corresponding to the two directions of the conditional involved (believe p iff p is true). On one hand, there is the goal of believing truths – believing p if p is true. And on the other hand, there is the goal of avoiding error – believing p only if p is true, or equivalently, not believing p if p is false. The pursuits of these two goals actively trade off against each other. If one has pretty good but defeasible evidence for p, one increases one's chances of believing the truth significantly by believing p, and increases one's chances of avoiding error significantly by suspending judgment. So, in the conscious decision about whether to terminate or continue inquiry, one inevitably consciously sacrifices one goal for the other.[278] It isn't enough simply to ask oneself whether one "finds the evidence for p convincing".[279] One faces the question of how convincing the evidence needs to be, and how much doubt it needs to foreclose, to believe.

Bernard Williams memorably wrote (apparently echoing Hume) that belief is "something that happens to us".[280] David Owens writes that "it is *the world* which determines

---

[277] James (1896).

[278] For development of this point, see (in addition to James himself), Foley (1993: 198-201); Owens (2000: 23-27); Ganson (2008: esp. 445-8).

[279] Cf. Hieronymi (2006: 51).

[280] Williams (1973: 148).

what (and whether) I believe, not *me*"; that once I decide to attend to the evidence, "the evidence takes over and I lose control."[281] These claims are overstated. If the world determined what we believed on its own, attendance to the same evidence would guarantee agreement in belief.[282] A picture on which belief just happens to us neglects our *pervasively* active role, across a wide range of cases, in setting the threshold of evidence that we require for belief, depending on our weighing of the two goals of believing truth and avoiding error.[283] A good explanation of why believing at will is so difficult will do justice to this ability of ours, as well as to the fact that this deciding how to weight these two goals is a necessary part of our doxastic lives that need not involve us in irrationality.

4.5 *An explanation in terms of coherence requirements?*

As I have said, the foregoing points were not supposed to show that the idea that belief aims at truth or that truth is the (constitutive) norm of belief could never be part of an explanation of why believing at will is so hard. Rather, it was supposed to show that even the best version of this idea is on its own incomplete as such an explanation.

---

[281] Owens (2000: 12), his italics.

[282] These quotations also neglect our role in interpreting our evidence and forming (potentially fallible) judgments about what it supports. Again, it is what we *take* our evidence to support that (direct) matters to what we end up believing, not what it actually *does* support. A picture on which evidence just somehow compels belief on its own misses this.

[283] It's odd that Owens seems to overlook this, since he himself nicely makes an analogous point in attacking an evidentialist theory of reasons: part of what settles whether one ought to believe must be the pragmatic considerations that determine where the threshold for sufficiency of evidence *ought* to be set. His view seems to be (cf. Owens 2000: 27, 50-1) that although such pragmatic consideration are normatively relevant, at the psychological level they cannot enter into one's deliberations about what to believe explicitly (cf. also Adler 2002b: 61-2). I just don't think Owens has the phenomenology right here. As I said above, it seems like I can often explicitly think to myself, "the costs of inquiring further are prohibitive, so I'll just accept this now," or "the costs of inquiring are very small, so I'll continue to investigate this before making up my mind." Of course, sometimes the influence is not explicit in reasoning – as with the influence of any consideration on one's attitudes. But it doesn't seem to me that it can't *sometimes* be explicit.

In offering my explanation, I want to in one way make do with less, and in another way make do with more. I want to make do with less in the sense that I will not be assuming that truth is the constitutive norm of belief in any particularly robust sense. That isn't to say that my explanation is incompatible with such an idea, just that it does not rely on it. I want to do with more in the sense that I am going to try to supplement my picture of the relationship of belief and truth with other claims that fill out the explanation of why believing at will is so difficult.

Crucial to my explanation is the requirement (ILC) that we encountered in chapter one, especially part (ii) of it:

Inter-level coherence (ILC). Rationality requires of S that

(i)   If S believes that her evidence supports D(p), then she takes D(p)

(ii)   If S believes that her evidence does not support D(p), then she does not take D(p)

Here, remember, D(p) is some possible doxastic attitude D towards some proposition p.

I will argue as follows: the paradigm cases of believing at will would have to involve violating (ILC) in a way that is transparent to oneself. But, as I argued in section 2.1, it is difficult, quite generally, to violate requirements of coherence like (ILC) in a way that is transparent to oneself. Thus, it is difficult to believe at will in the paradigm cases.[284] Moreover,

---

[284] There are some of forerunners of my explanation in the literature. Both Adler (2002b) and Owens (2000: esp. ch. 3) suggest at times that the difficulty of believing while also judging oneself to lack sufficient evidential grounds for belief may be important. But neither identify this as a *sui generis* coherence requirement, clearly distinguished from the substantive norms governing belief (indeed, Adler at least is clear that he takes the latter to be fundamental when it comes to explaining the difficulty of believing at will). Nor do they spell out the connection between the difficulty of violating this requirement and the difficulty of believing at will. Closer to my explanation is Setiya (2008). However, Setiya still does not present the explanation as one that makes

I will suggest that in the non-paradigmatic cases where believing at will is easier, what is different is that one does not have to violate (ILC) in a way that is transparent to oneself.

Before I begin with this argument, I want to engage with the grounds for (ILC). I defended this requirement in sections 1.3 and 1.5. But I want to engage with one particular thought about (ILC). One might think that what grounds (ILC) is itself the fact that truth is the norm of belief. If that's true, then my explanation of the difficulty of believing at will is ultimately parasitic on the claim that truth is the norm of belief. So my claim to have avoided relying on this will turn out to be false.

Why should one think that (ILC) is parasitic on the claim that truth is the norm of belief? The idea would seem to be that, since (as we noted in the previous section) a consideration in favor of believing p is evidential in nature iff it makes p more likely to be true, if one is committed to believing the truth, one is committed to believing what one takes one's evidence to support. So if the truth norm is constitutive of believing in such a way that in believing one must be trying to conform to the truth norm, one thereby will be irrational if one fails to satisfy (ILC). The truth norm explains *why* believing p whilst taking one's evidence not to support believing p should be incoherent.[285]

I agree that this argument give us *one* potential way of motivating (ILC). Some of the questions raised in the previous section still arise: why does it follow from it being constitutive

_____

coherence requirements as such fundamental. Rather, he presents it as a brute fact about belief (for which he can give no further explanation) that it isn't possible to believe in defiance of one's judgment about the evidence. In fact, as I'll make clear, I think this claim is too strong as it stands: as I'll argue, what is difficult is to do this *transparently to oneself*; it is not impossible to do it *at all*. Thus, on my view, but apparently not on Setiya's, one can still think of the relevant notion as a *requirement* that it is possible to violate, albeit not transparently. Relatedly, Setiya does not fit the explanation into a broader picture about how coherence requirements *more generally* are hard to transparently violate, which helps to make the account seem less brute and *ad hoc*. Finally, Setiya does not develop the explanation so as to show why the non-paradigmatic cases of believing at will, unlike the paradigmatic cases, are possible – a feature which I take to be crucial to my account.

[285] Cf. Adler (2002b); Shah (2003: 453).

of believing that one's belief is *subject* to the norm of truth, that one cannot flout such a norm without incoherence? But suppose those questions can be answered. In that case, I think that my explanation will turn out to be the right way to fill out the explanation of why believing at will is difficult. In that case, it still makes progress on the bare suggestion that truth is the norm of belief, by showing what requirements of coherence follow from that claim, and explaining the difficulty of believing at will in terms of the difficulty of violating those coherence requirements.

However, I also think that there is another way of motivating (ILC) without relying on the claim that truth is the norm of belief – and we encountered it in section 1.3. Instead of beginning with the idea that truth is the constitutive norm of belief in some especially robust sense, it starts with the point that the truth-predicate in belief attributions is redundant: that is to say, to believe p just is to believe that p is true. But, as I argued, it is incoherent to believe that p is true, while also believing that p is not likely to be true. The sense of likelihood at issue here turns out to be that of *likelihood on one's current evidence*. Since for a proposition to evidentially support a belief just is for it to be make the proposition believing more evidentially likely, we arrive at a justification of (ILC).

The line of reasoning just given begins with the redundancy of the truth-predicate. Interestingly, some defenders of the idea that truth is the constitutive norm of belief have explicitly argued that the redundancy of the truth-predicate cannot play an important role in explaining the difficulty of believing at will. The reason is that the truth-predicate seems to be redundant for other propositional attitudes: to suppose p is just to suppose that p is true; to assume p is just to assume that p is true; to pretend p is just to pretend that p is true.[286] Yet

---

[286] Shah (2003: 448); Velleman (2000: 247-50); Railton (1994: 74); Hieronymi (2006: 49).

these other attitudes are easy to take at will. So it might seem like the redundancy of the truth-predicate does no work here, and a more substantive norm of truth must be invoked.

I think this is too quick. For our argument begins with the redundancy of the truth-predicate, but will not generalize for other propositional attitudes. What is difficult about believing at will, I am claiming, is that it typically requires one to believe p while also believing that one lacks adequate evidence for p, and that is incoherent. Consider, by contrast, supposition. There is no similar difficulty with supposing at will because there is no coherence requirement forbidding one from supposing that p and *believing* that one lacks adequate evidence for p. Yes, the truth-predicate is redundant for supposing, and that gets us the claim that supposing p just is *supposing* that p is true. But whereas there is an argument to an incoherence between *believing* p is true and believing that one lacks adequate evidence for p, there is no argument of the same kind that gets us to an incoherence between supposing that p is true and believing that one lacks adequate evidence for p. So one can coherently suppose p despite believing that one lacks adequate evidence for p. So, supposing at will does not require one to be incoherent in this way. The redundancy of the truth-predicate needs supplementing with further claims about coherence to explain the difficulty of believing at will (as opposed to supposing at will), but it does not require us to invoke a substantive truth norm for belief.

As I have been stressing throughout this dissertation, it is important to distinguish coherence requirements on one hand from substantive reasons and norms on the other. This distinction reinforces the difference between the present explanation on the difficulty of believing at will on one hand, and an explanation in terms of the substantive norms for belief on the other. What one has substantive reason to believe depends on what one's actual evidence *actually* supports, not what one takes it to support. By contrast, (ILC) requires

coherence between one's first-order beliefs and one's higher-order *beliefs* about what one's evidence supports.

One can satisfy (ILC) while failing radically to believe what one's evidence actually supports. That one's beliefs about what one's evidence can support can diverge from what one's evidence actually supports is a quite obvious and apparently mundane epistemological claim. For example, to adapt an example that we encountered in sections 2.3-2.5, I might believe that the appearance of dew on the grass in my garden is decisive evidence that the magical garden fairy exists and has been crying. But (barring some *extremely* subjective view of evidence) this (higher-order) belief is of mine is just false: the dew on the grass is *not* decisive evidence (or indeed, somewhat more arguably, even *pro tanto* evidence[287]) that a magical garden fairy exists and has been crying. If I believe that a magical garden fairy exists and has been crying I will be believing against my evidence, despite my belief that my evidence supports this belief. I will not, however, be violating (ILC).

Although the possibility of a gap between what one's evidence actually supports and what one thinks one's evidence supports is apparently mundane, and is taken for granted for the purposes of most epistemology, it is a point that can be curiously elided in the literature on belief formation and believing at will. So one finds writers saying things like "believing in opposition to one's evidence is motivationally unintelligible,"[288] or "one particular belief-forming process, reasoning, is regulated solely by evidential considerations,"[289] or "belief aims

---

[287] Perhaps it's *very* weak *pro tanto* evidence; but there will surely be some examples where not even this is satisfied: suppose I believe that the fact that there's dew on the grass is evidence that the president of Sri Lanka is displeased. Or just pick your own obviously false belief about some *pro tanto* evidential support relation.

[288] Adler (2002a: 8).

[289] Shah (2003: 462). Also (*ibid.*: 469): "my deliberation won't count as belief-formation…unless the deliberation is solely influenced by evidence."

to 'track truth' in the sense that belief is subject to immediate revision in the face of changes in our all-things-considered evidence."[290] Interpreted charitably, these writers mean to say something like "to what one takes to be one's evidence", "putative evidence," and "perceived changes in our all-things-considered evidence," respectively. The reason it is so easy to slide into omitting these qualifiers, I think, is because when one is focused on distinguishing *evidential* considerations from *pragmatic* considerations, and on the contrasting role that each play psychologically in first-personal deliberation about what to believe, the considerations which one wrongly takes to be of evidential significance play just the same role as the considerations that one rightly takes to be of evidential significance – in contrast to pragmatic reasons. So "evidential reasons" gets used to actually *mean* something like "putatively evidential reasons", rather than to refer to *genuine* evidential reasons, in this context.[291]

Yet the slide is not innocuous: it misleadingly suggests what explains the difficulty of believing at will has to do with some difficulty with violating the substantive norms and reasons surrounding belief, rather than with violating coherence requirements. That, I am suggesting, is not right. After all, there is no puzzle about how I can believe that the magical garden fairy exists and has been crying in the case where I take the dew on the grass to be good evidence for that claim – even though, in fact, it isn't. What is hard to understand is how I can believe such a thing when I *take myself* to have lousy evidence for it. Some existing discussions have not kept the substantive norm of believing what the evidence supports and the coherence requirement (ILC) adequately distinct.[292]

---

[290] Gendler (2008b: 565).

[291] Cf. also Williams (1973: 141), whose uncharacteristically ungainly phrase "believing the evidence" slyly equivocates between possessing evidence and believing that one possesses it.

[292] See, e.g., Adler (2002b) and Feldman (2005).

The confusion here may be partially due to an ambiguity in talk of what it is to "consciously violate" a norm, an issue that we touched on in section 2.1 and that I will now expand on. On a weaker, less demanding notion of what it is to consciously violate a norm, it is to violate that norm without any failure of transparency of one's own mental states to oneself. On a stronger, more demanding notion of what it is to consciously violate a norm, it is to violate a norm in the full knowledge *that* one is violating a norm. The weaker and stronger notions of conscious violation of a norm are different because even if one knows what mental states one has, one may be mistaken about whether those mental states violate any norm. In the weaker case, one knows that one has the mental states that (as a matter of fact that one may not be aware of) violate the norm. In the stronger case, one knows not just that one has those states but that they violate the norm.

Now, all allow that one can violate the norm of believing what one's evidence supports in ways that involve a failure to be transparent to oneself – for example, in cases of wishful thinking.[293] However, some writers write as if that is the *only* way that one can violate the norm of believing what the evidence supports.[294] In other words, they write as if one cannot *consciously* violate this substantive norm even in the first, weaker sense. On this view, failures to believe what one's evidence supports are all ultimately down to some kind of failure to be transparent to oneself: to self-deception, or forgetting, or distraction, or fragmentation, or quasi-automatic, non-reflective belief.

---

[293] Pace Price (1973: 140), who claims one actually cannot *believe*, strictly speaking, without adequate evidence. This is surely too strong, unless one has an incredibly subjective view of evidence. As I'll suggest below, there are other ways of distinguishing belief from other attitudes that don't assume anything this strong.

[294] As I read him, this is Adler's view (Adler 2002b: 73-4, 78); Shah seems to edge toward a similar view at times (e.g. Shah 2003: 465, 473; Shah & Velleman 2005: 500). As I am about to argue, this is still too strong.

But this is just obviously not right, as shown by the case where I believe that the magical garden fairy exists and has been crying. In this case, my belief need not be due to any of these kinds of mental non-transparency. Rather, it is due to my false normative belief to the effect that the dew on the grass supports believing that the magical garden fairy exists and has been crying.[295] This case shows, then, that in the weaker sense of "consciously violate", it *is* possible to consciously violate the substantive norm of believing what the evidence supports.

By contrast, I claim, (ILC), like other coherence requirements (see section 2.1 above), is hard to consciously violate *even in this weaker sense*. That is: it is hard, without some kind of self-deception or failure of transparency, to even have the mental states that violate (ILC): to believe p, while also believing that one's belief in p is not supported by the evidence, and to recognize that one has all these beliefs. No thought or knowledge *that* these beliefs violate a coherence requirement is needed for the violation of that coherence requirement to be difficult to sustain. So (ILC) and the substantive norm of believing what the evidence supports are not on a par when it comes to the difficulty of violating them. (ILC) cannot be consciously violated even in the weaker sense, whereas the substantive norm can be.

But now notice that if it is difficult to consciously violate (ILC) in this weaker sense, it just *follows* that it is difficult to consciously violate the substantive norm of believing what the evidence supports in the *stronger*, more demanding sense of "consciously violate": that on

---

[295] Perhaps the failure to notice this simple possibility, and to want to put such failures down to failures of transparency, is due to the tacit assumption that what the evidence supports is (typically?) just obvious, such that this could not itself be the source of error. I disagree profoundly with this. As I have been stressing throughout this dissertation, it is a normative question whether a piece of evidence supports a possible doxastic attitude, and like other normative enquiries, figuring out what a detailed mass of cross-cutting pieces of evidence support believing overall can be genuinely *difficult*. Nor is the danger of believing out of accord with the evidence simply removed, as Adler and Shah seem to assume at times, by eliminating the sources of distraction and fragmentation, and just *deliberating* more. It is at least as common for those who believe out of accord with their evidence to have built-up, detailed rationalizations for their views which portray these views as supported by the evidence as it is for them to be simply ignoring something which it ultimately obvious to them. Turning to higher-order reflection is no panacea for bringing one's beliefs into accordance with the evidence.

which to consciously violate a norm is to violate it while knowing that one is violating it. For if one knew that one was violating the substantive evidence norm, one would know that one's belief was not supported by the evidence. But then, one would be in the position of believing whilst also believing that one's belief is not supported by the evidence; that is, of transparently violating (ILC). So to the extent that we grant that it is difficult to transparently violate (ILC), we also grant that it is difficult to consciously violate the substantive norm of believing what the evidence supports in the more demanding sense of "consciously violate".

We now have an explanation of what was left unexplained at the end of section 4.4: namely, why it is hard to consciously compromise on conformity to the substantive norms associated with belief. It is hard to do that *because* it would involve one in transparent incoherence. But that is not the case with all violation of substantive norms. There are many substantive norms that one can violate *without* being transparently incoherent, and those are easy to violate. By contrast, *all* genuine coherence requirements – I'll shortly be arguing – are hard to transparently violate. Whether a substantive norm can be consciously compromised on is a function of whether doing so would involve transparent incoherence. So it is coherence requirements (and the difficulty of transparently violating them) that are fundamental in the fullest and most illuminating explanation of the phenomena.

4.6 Explaining the cases

Suppose that I am right that it is difficult, quite generally, to violate coherence requirements in a way that is transparent to oneself. I now want to show how this fact – if it is a fact – does work in explaining why *believing at will* is difficult, in the paradigmatic cases – and why it is less so in the nonparadigmatic cases. If it can do all this, then the explanatory power of the

hypothesis that violating coherence requirements transparently is difficult will itself help to justify it.

Let's begin with the paradigm cases, like the one I began with where I offered you $1,000,000 to believe that the text is green. In this case, you no doubt believe that your evidence does not adequately support believing that the text is green. It might be an implicit belief, and not one that you have explicitly said to yourself aloud, but it is nevertheless a belief that you have. It is just obvious that your evidence doesn't support believing that the text is green. Holding fixed that higher-order belief, it follows that to come to believe that the text is green would involve coming to violate (ILC). But we need to say a bit more than this. After all, why must this higher-order belief be held fixed?

The answer can be given by recalling that we defined believing at will as believing something consciously and directly on the basis of pragmatic considerations. Pragmatic considerations are by definition non-evidential. So if you are taking something to be a pragmatic consideration, you take it *not* to bear on the evidential support for the first-order proposition. Thus, you cannot coherently think of it as *altering* the evidential support for that proposition, while regarding it as a pragmatic consideration. That is what precludes you from revising your higher-order belief (that the evidence does not support believing that the text is green) as part of your response for the monetary offer. If you are thinking of the $1,000,000 as a pragmatic consideration, then you take it not to bear on the evidential support for that proposition, then you will *still* take yourself to lack evidential support for that proposition; your previous belief will persist.[296]

---

[296] One might suggest that to fill this out, we need to posit a second rational requirement, in the form of some kind of diachronic constraint on belief persistence: rationality requires that if you thought you had lousy evidence for p, and you still remember having thought that, and you don't think you made a mistake previously, and you don't think anything has changed evidentially, then you still think you have lousy evidence for p. I'm

Of course, perhaps there is some complex and indirect method by which you could alter your higher-order belief.[297] But in such a case, you can *also* then correspondingly alter your first-order belief. That just is the case of plotting against yourself that I described earlier. It will be a case in which you take yourself at the time of forming the belief to be believing for evidential reasons, not for pragmatic reasons – and so it fails to meet our definition of believing at will.

The paradigm cases of believing at will are all cases where you are asked to form a belief in defiance of your higher-order judgment about what the evidence supports, violating (ILC). The paradigm cases are never ones in which you consciously believe that the evidence *does* support the belief. And there is good reason for this – for recall that according to part (i) of (ILC), it's also incoherent to judge that your evidence decisively supports a belief, but not to have that belief. So if you already do think your evidence supports some belief, then it would be transparently incoherent not to have that belief already, before pragmatic considerations come into play.

That leaves out, however, an intermediate case: that where you aren't sure about what the evidence supports, or where you think the evidence is itself indecisive between different potential attitudes. I'm now going to argue that this is one place where pragmatic considerations *can* in fact play a role. In fact, I'm going to suggest that several of the nonparadigmatic cases – where believing at will did seem possible – fall into this category. As

agnostic on whether we need this, or whether the descriptive claim that if you don't see yourself as having gained new evidence, then your higher-order belief thereby *won't* change, is enough.

[297] Although, there may be limits on how manipulable your higher-order beliefs will be. They themselves may be subject to constraints of generality. It is hard to see how you could coherently come to regard the evidence as supporting belief that *this* sentence is green, without regarding the evidence as supporting belief that other similar sentences nearby as green. Indeed, both higher-order beliefs are probably produced by a general disposition to treat appearances as evidence. That's why you would need some kind of general narrative like the one provided by GTCHA.

background to this, remember that, as I said (and defended) back in section 1.3, (ILC) does not forbid believing in a case where one isn't sure what one's evidence supports. With that in mind, let's consider the four nonparadigmatic cases I introduced:

1. *Some cases of religious belief.* As I said in section 1.3 above, there are numerous ways to understand this sort of case such that it involves no violation of (ILC). I propose that the right way to understand these cases involves one of the following possibilities:

    - The agent either *suspends judgment*, or lacks any attitude at all, at the higher-order level about what first-order attitude is required of her. In the case of religious belief (as contrasted with the case of what color the text is), it's very plausible that many of us are unsure what our evidence supports: we just think, "I have no idea what my evidence supports when it comes to this God business."

    - The agent believes that she is in a permissive case, where the evidence does not adjudicate between believing that God exists and suspending judgment.[298] So she doesn't believe that her evidence fails to support believing that God exists; nor does she believe that it decides it for certain.[299]

    Given either of these diagnoses, coming to believe that God exists on the basis of pragmatic reasons would not involve coming to violate (ILC), transparently or otherwise. That helps to explain why believing in God at will seems possible for some people. At the same time, those who have a determinate judgment that the evidence

---

[298] As noted in fn. 56 above, there's a debate about whether such cases are actually possible, but all participants in this debate should be able to agree that it's possible for agents to *take* themselves to be in permissive cases.

[299] Cf. James (1896). See also Bovens (1995: 824).

doesn't support believing in God will find it hard to come to believe in God at will, for example in response to Pascal's Wager. Our explanation of the difficulty of believing at will nicely separates these cases.

2. *Trusting the testimony of friends.* These cases may be explained in the same ways as the belief in God cases. Alternatively, perhaps what is going on here is that one is able to tacitly take one's friend's testimony to carry supreme evidential weight. It might be right here to diagnose one's state of mind as fragmented: at some level, one knows that one lacks sufficient evidence for believing that one's friend is innocent. But that isn't something that one transparently acknowledges, and so believing one's friend is not a *transparent* violation of (ILC).

3. *Believing in oneself.* In some of these cases, as noted, the act of believing itself alters one's evidential situation. So, it might be that once one does believe that one can win the game, it becomes quite likely (on one's evidence) that one can win the game. (ILC) again explains why believing at will in such a case is more possible: if one can, as one forms the belief that one can win, also form the belief that one's evidence supports believing that one can win, then once again one will not violate (ILC). Moreover, even if the act of believing has a fairly weak effect on one's evidential situation, if one oneself *overvalues* this impact, one will still avoid violation of (ILC) – despite violating the substantive norm of believing what one's evidence supports.

4. *Deciding whether to terminate or continue inquiry.* In these cases, I propose that what is going on is that one is able to exercise latitude over what one takes to be the bar for

something's being *sufficiently* supported by the evidence. (ILC) requires that one refrain from believing something that one takes not to be (adequately) supported by the evidence. But what is the bar for (not) being adequately supported, given that we want to believe truth but also avoid error, and that these two goals trade off against one another? The answer is: at least for the purposes of understanding the *judgment* of adequate support that features in (ILC), it is up to the individual believer, and it varies from circumstance to circumstance. The individual believer must refrain from believing that which she deems too unlikely on the evidence to warrant belief *in the circumstances*. Pragmatic considerations (such as the costs of future inquiry) may influence where she sets the bar for what is too unlikely for belief to be adequately supported, and so may her own personal dispositions in terms of how careful she is, how risk-averse, and so on.[300] Moreover, different people have different standards in how much evidential support they take to be required for believing (rather than suspending judgment) with respect to some particular subject-matter. Our explanation of the difficulty of believing at will in terms of (ILC) allows for this, as well as for the phenomenology of it seeming to you that you can choose as to whether to believe or to continue inquiry. Your decision as to where to set the bar for sufficient evidence determines whether you judge yourself to lack sufficient evidence for believing (in the circumstances). If you decide to set the bar low enough as to what counts as sufficient, you can then believe without violating (ILC). So you can decide to set the bar low, and

---

[300] Contemporary psychology bears out that this is consistent with our actual practices. See Nagel (2008: esp. 281-3).

thereby believe, or to set the bar high, and thereby not believe. So our explanation of the difficulty of believing at will also allows for this important kind of latitude.[301]

In summary, the nonparadigmatic cases where one can believe at will are possible because all of them fall short of transparently violating (ILC). Some are cases where one does violate (ILC), but non-transparently. Others are cases where one does not violate (ILC) at all, because one has lacks a higher-order judgment about what one's evidence supports, or represents that evidence as permissive, or can decide for the purposes of that judgment what counts as sufficient evidence. But none are transparent violations of (ILC). It is a major virtue of the present explanation that it can explain why these cases are possible, where the paradigmatic cases of believing at will are not. The account in terms of substantive norms, as we saw, cannot do that.

The present account can also shed light on the nearby cases discussed in section 4.2 that do not count as believing at will at all. In cases of wishful thinking, one may violate (ILC), but not in a way that is transparent to oneself. In cases of plotting against oneself, one never at any *single* point in time takes oneself to be violating (ILC). Indeed, (ILC) nicely explains why instances of plotting against oneself often work via manipulating one's higher-order beliefs about one's evidence. For example, one starts going to GTCHA so that one will come to regard the belief that the text is green as supported by the evidence (or at least, stop believing

---

[301] The point here is not just that it is sometimes indeterminate or even indeterminate-seeming what the evidence supports (cf. Adler 2002b: 63). Whatever the evidential state of play is, we face the question of how much support to require for belief. Adler says that we "minimize" the risk of error "to the extent that belief is in accord with the evidence." But this elides the issue: we would minimize the potential risk of error by being maximally cautious and withholding belief from anything not completely certain; we would maximize the potential reward of true belief by just believing everything that is even minimally probable. In between these extremes, we have to figure out what bar to set, and just saying "believe in accord with the evidence" doesn't help.

that it isn't). These cases, too, are possible because they do not involve transparent violation of (ILC).

4.7 The difficulty of transparent incoherence

In section 2.1, I posited that it is difficult, in general, to violate coherence requirements in a way that is transparent to oneself. I want to say a bit more to justify this in the context of (ILC) here.

Some philosophers have suggested, or come close to suggesting, something stronger than this: that the kind of incoherence involved in violating (ILC) is actually impossible *simpliciter*.[302] I think this overstates things. We are capable of tolerating a certain amount of internal incoherence in our mental lives. But importantly, to sustain the incoherence we have to put the explicit recognition of the incoherence out of our minds to some extent. That is, we can be incoherent, but only by failing to be fully transparent to ourselves.

For example, suppose that Fabian considers himself to be extremely attractive to most members of the opposite sex. Suppose also that Fabian is aware of a body of psychological research that shows that people like him tend to systematically overestimate their attractiveness to the opposite sex, and that the women he tries to seduce often ask him to leave them alone. When Fabian reflects on all of this, he is inclined to admit that his evidence that he is extremely attractive to most members of the opposite sex is pretty lousy. But he doesn't like to dwell on that. When he starts to think like that, he just jumps in his sports car, rolls down the windows, turns the volume on his stereo up to 11 and goes for a spin, and very soon he stops thinking

---

[302] Hampshire (1965); Pettit & Smith (1996: 448); Setiya (2008: 43). Adler (2002a, 2002b), Hurley (1989: 130-5, 159-70, and Wallace (2001: 12-13) all also seem tempted by this view; some of them ultimately qualify it in some way.

about it. His belief that he is extremely attractive to most members of the opposite sex survives.

Now, one could argue that there is never a single moment where Fabian both believes that he is extremely attractive to most members of the opposite sex and believes that this belief of his not well-supported by the evidence. On this diagnosis, in his moments of reflection his belief that he is extremely attractive is suspended, so that he no longer counts as believing, while at all other times he does not count as believing that his belief is well-supported by the evidence. While we can perhaps imagine the case that way, I also think that saying that this is the only way of making sense of the case is *ad hoc*. Most of the beliefs that we have at any particular point in time are not occurrent at that exact moment; we can put many of our beliefs out of our minds while still counting as believing them. I believe that when I was nine years old a very embarrassing incident occurred at my birthday party, but fortunately I manage to put it out of my mind most of the time. When I do so, I don't count thereby as suspending judgment about whether the incident occurred. I don't see why Fabian can't be the same. Most of the time, he manages to put the meagre evidential basis for his belief in his own attractiveness out of his mind. He doesn't thereby cease to count as believing that the evidential basis for his belief is meagre.[303] Meanwhile, his first-order belief that he is extremely attractive continues to play a role in explaining his behavior, in a way that makes it right to attribute that belief to him.

Nevertheless, what the need to tell the story of Fabian in such a way that his mental states are non-transparent to him illustrates is that it is difficult to violate coherence

---

[303] Cf. Elga's (2005) diagnosis of his own state of mind with respect to overrating his own abilities.

requirements like (ILC) in a way that is *transparent* to oneself.[304] This, I think, is a more general feature of coherence requirements. I already gave one example in section 2.1: that of the instrumental requirement. As I said, it's important to our ability to make sense of cases of instrumental irrationality that we posit some kind of mental non-transparency to the agent in question.

Here is another example: the requirement of transitivity on preferences, which requires that one not (prefer A to B, prefer B to C, and prefer C to A).

Here is a case of violation of transitivity which is easy to imagine. Consider the three following things that a philosopher might do with his Saturday: working on his new article, volunteering at the homeless shelter, or re-watching series 4 of *Friday Night Lights*.

- Attending to the options of working on his article and volunteering at the homeless shelter, working on the article seems like an important project that he can justifiably pick over volunteering, and which allows him to stay in his pajamas and not have to interact with anyone. So he prefers working on his article to volunteering.

- Attending to the options of volunteering at the homeless shelter and re-watching series 4 of *Friday Night Lights*, choosing to do something so trivial as watch TV rather than volunteering seems callous. So he prefer volunteering to watching TV.

---

[304] Note that the claim is not that it is impossible to, without non-transparency, believe something which one judges not to be demonstrated, proven or "shown" beyond doubt by one's evidence (cf. Hieronymi 2009: 181). Nor is the claim that one cannot believe something about which one lacks any definitive higher-order judgment regarding what the evidence supports. The claim is just that one cannot, without non-transparency, believe that which one believes one's evidence does not *adequately* support believing. More on what adequate support is below.

- But, attending to the options of watching *Friday Night Lights* and working on his new article, the writing of the article seems difficult and energy-consuming after his long week. So he prefers watching TV to working on his article.

These preferences are intransitive. Yet what seems much harder to imagine is the philosopher having all three options vividly before his mind, and sincerely declaring, "I prefer working on my article to volunteering, I prefer volunteering to watching TV, and I prefer watching TV to working on my article." Or, if we can imagine such a claim, it is as a sort of joke. Once the philosopher vividly attends to the intransitivity, he will feel a pressure to resolve it. Yet he can get away with never vividly attending to it. Here's one way that it's depressingly likely to go: by focusing first on the choice between the article and the volunteering, he rules out the volunteering and puts that out of his mind. Then he compares the article and the TV, and picks the TV. So, he ends up watching TV, never attending to the comparison between volunteering and watching TV.

Once again, one could conceivably claim that the philosopher therefore doesn't *really* count as violating transitivity (construed as a synchronic requirement) at all. One could claim that his preferences simply change, depending on which pair of options he attends to: there is no single point in time that he violates transitivity. But this way of thinking of things again neglects the importance of *dispositions* in ascribing attitudes. Economists and decision theorists think of preferences between two options as constituted by what would hypothetically be chosen given a pairwise choice between those options. And at one *single* point in time, the philosopher can be disposed such that, *were* he offered a choice between working on his article and volunteering, he'd pick the article; *were* he offered a choice between volunteering and watching TV, he'd pick volunteering; and *were* he offered a choice between watching TV and

262

working on his article, he'd pick watching TV. Insisting that he only counts as having these preferences when the offers are actually made severely compromises our ability to represent him as having (relatively) complete sets of preferences over possible options, in the real world where actual strict pairwise choices between options are rare. Moreover, such a move robs the theory of rationality of its bite. Represented this way, the agent is not guilty of intransitivity. But the example I gave is exactly the sort of case those who propose transitivity as a rational requirement want to count as irrational. The nice compromise verdict is that violation of transitivity is not itself impossible; what is at least very difficult, is violating transitivity in a way that is *transparent to oneself*.

As I also said in section 2.1, it's important to realize that there are many ways for your violation of coherence requirements not to be transparent to you. So the claim is not that coherence requirements cannot be violated at all. On the contrary: such violations may be quite widespread – so is non-transparency to oneself.

Is there something further that explains the general fact about the difficulty of transparently violating coherence requirements? It may be that we can get somewhat further in explaining why it is hard to violate coherence requirements by thinking of the nature of the attitudes involved. I suggested in the case of the instrumental requirement in section 2.1 that one simply doesn't *count* as genuinely intending an end if one consciously has that intention in view, and consciously refuses to intend the (consciously) believed necessary means to that end. Perhaps this can be generalized to other attitudes. If some mental state of mine isn't such that I am disposed to regulate it according to my higher-order beliefs about my evidence under conditions of full transparency, then perhaps that state just doesn't count as belief, and it

263

wouldn't be right to attribute belief; instead, it is some other state that falls short of belief.[305]

In fact, this strategy nicely points to a way of distinguishing belief from certain weaker cognitive attitudes *without* ruling out the nonparadigmatic cases of believing at will from counting as genuine beliefs.

This is an interesting proposal, and effectively amounts to the claim that the coherence requirements associated with belief are constitutive of believing. This makes progress on the claim that the relevant *substantive norms* are constitutive of believing. For as we saw, it cannot be right that a mental state isn't belief unless it *actually* tracks the truth – and the claim that a mental state isn't belief unless it's normatively evaluable against the standard of truth does not explain why it's hard to consciously violate this standard. In the case of coherence requirements, it's much more plausible to say not just that sets of beliefs can be evaluated for their satisfaction of these requirements, but that genuine beliefs will actually (when everything is transparent to the agent) generally satisfy them. That reinforces the importance of keeping substantive norms and coherence requirements apart.

---

[305] For example, an "alief" (cf. Gendler 2008a, 2008b; see also Setiya 2008: 43; Velleman 2000: 255-81; Millgram 1997: 31; Hampshire 1965: 78-9). So, for example, I know that my evidence does not support believing that the plane will crash; but still I feel terrified (to an extent as if it is going to crash); cf. Shah & Velleman 2005: 507. On the "alief" characterization, this does not amount to believing that the plane will crash, but rather alieving that it will crash – having the motivational dispositions as if one believed without certain crucial cognitive markers of belief proper. (Similar treatments might be given of cases of knowing overestimation of one's own abilities (cf. Elga 2005) or of knowing underestimation of the abilities of those from marginalized groups (cf. Gendler 2011). As Gendler presents things, one of the things that distinguishes belief and alief – and that makes it appropriate to attribute the latter rather than the former cases – is that belief is regulated by the *substantive* evidence norm (Gendler 2008b: 565). Like the explanation of the difficulty of believing at will in terms of belief's aiming at truth, I think this locates the relevant contrast at the wrong level. Even if one likes the belief/alief contrast (as I do), to rule out something's being a belief on the basis that it is recalcitrant in the face of changes in *actual* evidence is too stingy: it leaves too little counting as belief, and precludes the possibility of genuine beliefs that are systematically evidence-unresponsive. What Gendler should instead say, in my view, is that for something to be belief it has to be responsive (at least in conditions where one's mental states are transparent to oneself) to one's *judgments* or higher-order beliefs about one's evidence – that it, that is be governed by (ILC). This allows for beliefs that are systematically out of accord with the evidence, while simultaneously leaving a good range of cases, like the fear of flying case and (certain versions of) the implicit bias case, on the "alief" side of the divide.

Alternatively (or additionally), one might appeal to what it is for something to be a *coherence requirement*. As I suggested in section 2.1, it may be that what it is for mental states to be jointly incoherent just is for them to be hard to transparently and reflectively sustain, in full consciousness. Again, then, it will simply follow from (ILC)'s being a coherence requirement, given what a coherence requirement is, that it is hard to violate (ILC) consciously.

Ultimately, however, I want to suggest that at least in some sense our explanation simply bottoms out in the claim that violating (ILC) is incoherent, and as rational beings we find conscious incoherence hard to sustain (as evidenced by the wide range of examples of coherence requirements that are hard to consciously violate). The fundamental thing that explains the difficulty of violating (ILC) just is our rationality, and the explanation bottoms out there. I'll have more to say about this in the conclusion to the dissertation as a whole.

Chapter 5
Coherence and Deductive Constraints on Belief

5.1 Introduction

One might think that, on a coherentist account of rational belief, deductive constraints would be central. After all, when the word 'coherence' is mentioned, one of the first things that springs to mind is consistency. Some even use the terms interchangeably. However, there are well-known objections to deductive constraints on belief. One question one might have is whether these objections can be made to work in a framework that distinguishes, as I have, rational requirements on one hand, and reasons and evidence on the other. For example, one might think that in certain cases one's *evidence* supports having inconsistent beliefs, but that nevertheless, there is a rational requirement of coherence forbidding such a combination of beliefs.[306] In that case, we would just have another example of a case in which reasons and rationality come into conflict, as I argued that they do in other ways in chapter 1.

In this chapter, however, I will argue that the challenge to deductive constraints can be motivated internally within a coherentist framework. In so doing, I also argue that this particular way of generating the challenge is actually in many ways more robust than existing ways of generating it. In this way, we see another way in which distinguishing rational requirements and reasons sheds light on a central debate in epistemology. Consequently, I suggest that deductive closure (and, ultimately, consistency) requirements should be rejected, even within the coherentist framework. This leaves open the possibility that there are other weaker requirements of rationality that could in a broad sense be thought of as deductive; I explore that question in the final section of the chapter.

---

[306] Fogal (ms.) tentatively suggests this, without fully endorsing it.

5.2 Deductive closure and the preface paradox challenge

That deductive closure constraint that we will be considering for much of this chapter can be formulated as follows:

> **Deductive closure constraint.** Rationality requires of S that if S's total belief set logically entails some proposition p, then S believes p.[307]

Various worries might be raised about this deductive closure constraint being too demanding.[308] Still, even those live to such worries might maintain that obeying the deductive closure constraint is at least an *ideal*.[309] Or they might maintain that having beliefs that are closed under logical entailment is at least *permitted*:[310]

> **Deductive closure permission.** If S's current total belief set is rationally permissible, and S's total belief set logically entails some proposition p, then it is rationally permissible for S to believe p (while also retaining all the beliefs in her current total belief set).[311]

---

[307] NB: in this chapter I use the lower-case 'p' for propositions, and the upper-case 'P' for sentences that pick out those propositions.

[308] See, e.g., Harman (1986). Harman's first criticism is of a narrow-scope construal of this principle (cf. Harman (1986: 11; he himself does not put things this way). But he also targets a wide-scope construal of it (*ibid.*: 12).

[309] See, e.g., Kaplan (1996: 113).

[310] Thanks to Tamar Gendler for making this suggestion to me.

[311] This is a more general version of what sometimes gets called the "conjunction rule" (cf. Stalnaker 1984: 93; Foley 1993: 166), on which if it's rational to (believe p and believe q), then it's rational to believe (p & q). I take it that the use of 'rational' here usually and naturally designates 'rationally permissible'.

This captures the rough idea that it is permissible to extend one's belief set by performing logical operations on one's existing beliefs.[312]

So-called "preface paradox" cases, however, question even whether these weaker claims are true. They purport to show that a subject can be *rationally required* to withhold belief from (or even believe the negation of) a recognized logical consequence of her other beliefs, even if those other beliefs are themselves rational. Thus, they suggest that the deductive closure constraint is not even an ideal, and that the deductive closure permission is false.

The original preface paradox cases feature books and prefaces, and take a form where an author has written a long book containing a large number of claims – yet indicates in the preface of the book that the book no doubt contains errors.[313] This conjunction of claims – those in the body of the book plus that in the preface – appears to be inconsistent. Yet it seems that, even if every claim is the book is responsible and rational taken by itself, the author is right to think that there are probably errors – and to withhold belief from the claim that the book contains no errors. Moreover, to believe that the book contains no errors seems

---

[312] You will notice that this formulation is not wide-scope in the straightforward way that the deductive closure constraint is. One could formulate the deductive closure permission as the claim that, where some proposition p is a logical consequence of some set of beliefs SB, rationality *either* permits S to believe p (and/)or permits her to revise something in SB such that the new set no longer entails p. This is wide-scope and brings the deductive closure permission into closer analogy with the deductive closure constraint. But I worry that it is too weak. If rationality is generally quite permissive, it may often permit all sorts of revisions in one's current belief set, and so it could as a result be the case that for any individual proposition, it's always permitted to revise one's belief set such that it doesn't entail that proposition. That would make this proposed wide-scope formulation of the deductive closure permission true, but not in a way that would reflect anything general about the permissibility of drawing logical inferences from one's belief set. The deductive closure permission as I have stated it captures this idea more directly – while also avoiding the implausibly *strong* claim that for *any* total belief set (however crazy), if S's total belief set logically entails some proposition p, then rationality permits S to believe p. (Note that this claim would have the crazy result that by having inconsistent beliefs, I can make any belief rationally permissible for me, since the inconsistent set entails every proposition.)

[313] The preface paradox in its original form was formulated by D.C. Mackinson (1965).

irrational, since even though each individual claim has only a small chance of being false, these risks of error aggregate to make it very likely that there is at least one error in the book.

Unfortunately, although this presentation of the challenge to the deductive closure constraint from risk aggregation is nicely concrete and vivid, the details of the case introduce at least two distracting features that are in fact immaterial to the central point being made. First, as a few philosophers have pointed out,[314] strictly speaking the claims in the book plus the preface claim do *not* actually form an inconsistent set. Suppose that, in the actual world, the author writes some particular book consisting of the set of propositions $\{p_1, p_2, \ldots, p_n\}$. Call the proposition that there is at least one error in the author's book *e*. Now consider some other possible world in which the set $\{p_1, p_2, \ldots, p_n\}$ contains only true propositions, but the author writes a slightly different book, identical to that which she writes at the actual world, except containing $\neg p_2$ instead of p_2. That is a world in which the set $\{p_1, p_2, \ldots, p_n, e\}$ contains only true propositions. So the set $\{p_1, p_2, \ldots, p_n, e\}$ is consistent. In order to make it inconsistent, we have to add to this set the proposition that the author's book contains the propositions $\{p_1, p_2, \ldots, p_n\}$. That does the trick of making the set inconsistent. But when we return to our preface case, it is not plausible that the author has a correct single belief about exactly which propositions her book contains, or is capable of forming such a belief.

Second, one may doubt whether the writers of books really do (or should) believe every proposition that their books contain.[315] Belief-ascription is a tricky matter, and the mere fact that someone wrote a book saying something does not establish that she (outright) believes the content of what is said, even if she is being sincere in making the claim. So perhaps

---

[314] See Evnine (1999); Wedgwood (2007); Roush (2010).

[315] See, e.g., Stalnaker (1984: 93-94); Weatherson (2005: 429-31); Leitgeb (2014a). Leitgeb, for example, proposes interpreting authors of books as merely asserting that the claims are highly probable, and that taken as a set they are mostly true.

the author does not really believe every proposition in her book; again, this is another hurdle to representing the preface case as involving the author in a genuine inconsistency.

One could try to fix up the preface case to deal with these problems. However, I think it is less distracting to abstract away from the particular case, and simply note the features that will generate the problem for the deductive closure constraint (and its weaker cousins), since they are highly general and do not require any stories about prefaces or books. The feature of central importance to generating a case which puts pressure on the deductive closure constraint is that there must be a number of individual beliefs which each are in some sense highly likely but with a small chance of error. As one puts more and more of these beliefs together, the risk of error aggregates, until it becomes overwhelmingly likely that at least one of them is mistaken. Then, the thought goes, it will be irrational to believe some logical consequence of the large set, if that logical consequence is true only if the whole large set is true. The simplest example is a long conjunctive proposition formed out of the individual beliefs in the set, but other examples are possible too.

5.3 Ways of formulating the challenge to closure

The appeal just made to the possibility of error, and its aggregation, is vague. What is supposed to yield this possibility of error? One answer is that the subject's *evidence* leaves open the possibility of error. Here the claim is that, since one's evidence does not generally assign an evidential probability of 1 to the individual propositions which one believes, and evidential probability obeys probabilistic constraints, one's evidence assigns a very low evidential probability to long conjunctions of these beliefs. Then, on the assumption that rational

requirements cannot come into conflict with believing what the evidence supports, one can plausibly generate failure of the deductive closure constraint.[316]

As noted in the introduction to this chapter, this is not obviously a good objection to deductive closure construed as a putative coherence requirement of rationality. We already saw in chapter 1 that (coherence) requirements of rationality can sometimes conflict with what your evidence gives you most reason to believe. A defender of the deductive closure constraint might allege that preface paradox-type cases are also cases of this sort. On this view, showing that in preface paradox cases, fulfilling the deductive closure constraint would lead to a failure to believe what one's evidence supports, does not suffice for showing that the deductive closure constraint is false. For such cases may be further instances of conflict between coherence requirements and evidence-responsiveness.

Moreover, when formulated in terms of evidence, the challenge to deductive closure rests upon a particular view of evidence that is not uncontroversial. Most prominently, Timothy Williamson claims that your knowledge has epistemic probability 1 for you.[317] On this view, if you really know every one of a large number of propositions, there is no aggregation of risk in terms of evidential probability when it comes to their conjunction.[318] Of course, you are very unlikely to know every proposition that you believe, so it might seem that

---

[316] So, Easwaran & Fitelson (2015: 64-73) argue that deductive closure and consistency requirements conflict with the norm enjoining one to believe what one's evidence supports, and so must be rejected.

[317] Williamson (2000). See Williamson (2009) for application of this idea to preface-style cases. And see Smith (2010) for another account of evidential probability that is friendly to deductive closure constraints.

[318] Williamson (2009) tries to explain away the counterintuitive sound of this by pointing out that even though you know any conjunction built out of conjuncts that you know (and such a conjunction has probability 1), you may fail to know that you know many of the conjuncts, in which case the evidential probability *that you know* the conjunction will be very low, which is why it seems to you that you don't know the conjunction.

this strategy fails to solve the problem for the deductive closure constraint at least as stated.[319] But in Williamson's "knowledge-first" epistemology, rational belief is to be understood in terms of knowledge, such that a rationally ideal subject will believe only that which she knows – and this preserves the thought that a rationally ideal subject will obey the deductive closure constraint even as originally stated.[320]

An alternative way of generating the challenge to deductive closure from preface-style cases appeals not to the possibility of error given one's evidence but to the possibility of error given one's *credence* (graded belief). The crudest way to do this is to claim that outright belief is reducible to credence. For example, suppose that one accepts the so-called "Lockean thesis", according to which there is some threshold T such that for any proposition p, S believes p iff S's credence for p $\geq$ T.[321] Then, in preface-type cases, one will have to violate the deductive closure constraint if one is to remain probabilistically coherent (that is, if one's credences are to obey the laws of probability).

Unfortunately, this "Lockean" view is controversial (and, it seems to me, becoming less and less popular).[322] Many think that there is more to outright belief than just meeting some appropriate threshold of confidence, such as being disposed to act on, or assert, or reason with, the proposition in question. Some philosophers seem to think that by adopting

---

[319] Even if this were right, it might vindicate a version of the constraint that requires one to believe the consequences of that which one *knows*. Combined with Williamson's theory of evidential probability, this revised claim would escape a version of the paradox generated through the notion of evidential probability. But it would not escape my version of the paradox.

[320] At least to the extent that Williamson talks of 'rationality' at all. See section 0.7.

[321] Defenders of this view include Foley (1993), Sturgeon (2008), and more equivocally, perhaps Christensen (2004).

[322] For criticisms and alternative views, see amongst many others Kaplan (1996: sec. 3.3), Fantl & McGrath (2009: ch. 5), Buchak (2014), and Ross & Schroeder (2014).

such a view, they can avoid the challenge to deductive closure from preface paradox-type cases. If no arbitrarily high credence guarantees outright belief, it is thought, then there is no obvious tension between probabilistic coherence between one's credences on the one hand, and deductive closure between one's outright beliefs on the other.[323]

However, this is too hasty. The challenge to deductive closure can still be mounted without assuming the reducibility of belief to credence. We can take a step towards a less controversial version of the challenge by replacing the *descriptive*, reduction-generating claim with a *normative* claim about the *rational* relationship between belief and credence. This is essentially the tack taken by James Hawthorne & Luc Bovens (1999), who modify the Lockean thesis so that it says that an *ideally rational* subject will have some threshold T such that for any proposition p, she believes p iff her credence for p $\geq$ T. This avoids any reduction of belief to credence, since Hawthorne & Bovens nevertheless allow that many subjects will not satisfy this requirement. Their paper shows neatly that if one accepts the normative variant of the Lockean thesis, one can generate rational failures of deductive closure.

However, in order for this demonstration to work, it's crucial that they assume that an ideal subject has a *single* threshold for believing which is fixed regardless of context or subject-matter. That claim still very controversial; indeed, I think it is false. It is too demanding to say that subjects must have the same belief-threshold for all propositions regardless of subject-matter. For example, consider lottery cases. It seems that the credence-threshold we set for *my ticket will lose the lottery* is *much* higher than it is for most ordinary propositions,[324] and this is not obviously at all irrational. Moreover, Hawthorne & Bovens themselves express sympathy for

---

[323] See, e.g., Stalnaker (1984); Weatherson (2005).

[324] DeRose (forthcoming: ch. 11) brings this out nicely.

a pragmatic view of belief on which "belief might well be just a matter of [one] stipulating a confidence level that [one] finds high enough to be of special significance" (Hawthorne & Bovens 1999: 256). But surely what is of special significance might vary depending on context and subject-matter.

We should look, then, for a way of generating the challenge with even weaker assumptions about the rational relationship between belief and credence – and ideally from *within* the set of coherence requirements of rationality. This is what I will try to provide next.

5.4 A new way of formulating the challenge

What we need to generate this challenge are the following claims:

Weak fallibilism about belief. It's possible that

(a) There are many (suitably independent) propositions such that for each proposition p in this set

    i.   S does not have credence 1 in p

    ii.   S (outright) believes p

(b) S is not radically probabilistically incoherent

(c) In satisfying (a) and (b), S violates no requirement of rationality

Belief-credence coherence (min). Rationality requires of S that, if S has a credence below 0.5 in p, then S does not (outright) believe p.

The first principle is labelled as a kind of *fallibilism* about belief because it affirms the rational permissibility of believing a non-trivial number of non-trivially independent propositions,

without investing absolute certainty in these propositions (and without being radically probabilistically incoherent). It is a *weak* form of fallibilism because it makes no commitments about what is required (e.g. regarding the evidential situation, or error-possibility-space) to know the propositions in question, and it makes no commitment about the felicity of one's explicitly acknowledging the chance of error about one's beliefs in conversation.[325] Henceforth, I'll just call it 'weak fallibilism'.

Careful readers will notice that weak fallibilism is somewhat vague, in its usage of the words 'many', 'suitably', and 'high'. What is required, more precisely, is that there be enough propositions, that they be independent enough, and that S's credences for them be such that, in order to have a credence above 0.5 in their conjunction, S would have to be radically probabilistically incoherent.[326] That may sound somewhat demanding, but it is not. For example, if one has credence 0.9 in each of just 7 probabilistically independent propositions, the probabilistically coherent credence for their conjunction is 0.48; if one has credence 0.9 in each of 25 probabilistically independent propositions, the probabilistically coherent credence for their conjunction is 0.07. If one has credence 0.99 in each of 70 probabilistically independent propositions, the probabilistically coherent credence for their conjunction is 0.49; if one has credence 0.99 in each of 250 probabilistically independent propositions, the probabilistically coherent credence for their conjunction is 0.08.[327]

---

[325] I defend a stronger form of fallibilism about knowledge in Worsnip (forthcoming-a).

[326] I cannot put an exact characterization on 'radically', which is also vague. However, this ought not be a problem. For a start, those who think that perfect probabilistic coherence is permissible will immediately agree, *a fortiori*, that some non-radically-probabilistically-incoherent states are permissible. The only reason I can think of for denying that perfect probabilistic coherence is *permissible* is that some deny that it is permissible to be maximally confident in the logical truths (cf. Christensen 2007). But presumably even on this view, it's permissible to be *very* confident of the logical truths, which would prevent *radical* deviations from probabilistic coherence, on any reasonable construal of 'radical'.

[327] Ram Neta objected to me that the assumption of probabilistic independence here is questionable, especially as we tend to rely on just a few sources (memory, testimony, perception) for many of our beliefs. While it's true

275

Belief-credence coherence (min) is a wide-scope coherence requirement, like others examined earlier in this dissertation. So, en route to attacking deductive closure, I am positively proposing another coherence requirement, which I will defend in section 5.6. Belief-credence coherence (min) is labelled '(min)' because it effectively imposes a *minimum* lower bound, 0.5, such that it is irrational to simultaneously believe a proposition and yet also have a credence below the lower bound in that proposition. However, it is crucial to see that this is not a reduction of belief to credence in any sense. The idea is not that credence higher than 0.5 automatically counts as outright belief, nor is it even that it is always rationally permissible to believe something in which one has a credence (or even a rational credence) of higher than 0.5. Rather, it is just that it is irrational – because incoherent – to believe something that one has a credence lower than 0.5 in. Moreover, the by-now familiar points about wide-scope requirements apply to belief-credence coherence. That is to say, it does not give any special authority to your credences, saying that your credence determines what it is rational for you to believe. It does not say that, in some situation where you violate belief-credence coherence, it is your outright belief rather than your credence that should be revised. It just says that, one way or another, you should line the two up such that you do not combine a belief in p with a credence lower than 0.5 in p. For now, I'll drop the 'min' from the name, but we'll encounter an analogue 'max' principle later.

---

that the assumption of perfect probabilistic coherence is an unrealistic simplifying assumption, propositions can still be relatively probabilistically independent while being based on the same broad sources. This is so especially if most of the credence space that one gives to the possibility that one is mistaken about some proposition involves local error, where one's memory (or testimony-source, perception, etc) fails on this particular occasion but not in some more global way. Here it actually does help to return to the original preface case. Even though many of the propositions in the history book may be based on the same (coarsely-individuated) sources, they enjoy enough probabilistic independence that the probability of one of them being mistaken is high.

Now I'll spell out why weak fallibilism and belief-credence coherence require rejection of the deductive closure constraint that we began with. Suppose we have a doxastic subject, Gina. Suppose that Gina (outright) believes a large number of propositions, and assigns each one a very high, but non-zero, credence. Suppose further that Gina is not radically probabilistically incoherent. Given suitable independence between the propositions she believes, Gina will, if she is not radically probabilistically incoherent, assign the conjunction of these propositions – call it C – a credence below 0.5. According to weak fallibilism, given these facts so far, it's possible that Gina violates no requirement of rationality.

But if both belief-credence coherence and the deductive closure constraint were true, the facts specified so far would guarantee that Gina violates a requirement of rationality, for the following reason. Either Gina believes C or she doesn't. If Gina does believe C, she violates belief-credence coherence, since she has a credence below 0.5 in C. If Gina doesn't believe C, she violates the deductive closure constraint, since her beliefs entail C. So, by cases, Gina violates a requirement of rationality. That contradicts what weak fallibilism said about Gina's case. So weak fallibilism and belief-credence coherence require the rejection of the deductive closure constraint.

Neither weak fallibilism nor belief-credence coherence involves any commitment to reducing outright belief to credence. What this shows is that we can get a serious problem for the deductive closure constraint going without assuming anything about the reducibility of outright belief to credence. Nor do weak fallibilism or belief-credence coherence make the same problematic assumptions that Hawthorne & Bovens did. While belief-credence coherence imposes a requirement that a rational subject *not* believe something for which she has credence < 0.5, it does not impose any threshold such that a rational subject must *believe* a proposition that she assigns a credence ≥ that threshold. Nor does it require that a subject

277

have any such fixed personal threshold. This allows that a rational subject could have different thresholds for outright belief depending on context or subject-matter – as long as the threshold never dips below 0.5 – or indeed no specific threshold at all which suffices for her believing, even in particular contexts.

In the next two sections, I will defend weak fallibilism and belief-credence coherence, in turn.

5.5 Defending weak fallibilism

Weak fallibilism is a bare possibility claim. It reflects the *possible permissibility* (not even the requirement) of both (a) displaying an extremely minimal humility about many of your beliefs, by not assigning them credence one, and (b) not being radically probabilistically incoherent.[328]

There is one view about the relationship of belief and credence that can block weak fallibilism, namely the view that belief *is* just credence one. If that is right, then *a fortiori* one cannot permissibly have outright beliefs in propositions that one has non-one credence in (since one cannot have such outright beliefs at all). So weak fallibilism would be false. Interestingly, this shows that the one view about the relationship between belief and credence that can evade the problem for deductive closure as I have set it up is itself a reductive view, albeit a very special kind of reductive view. This is just the opposite of what is often assumed (cf. section 5.3 above). Even if it turns out that the view that belief is credence 1 is worth taking seriously, it would be an interesting result that one has to adopt this view in order to save the deductive closure constraint. Historically, I do not think that most of its defenders

---

[328] This leaves it open that rationality might require precise probabilistic coherence, or more weakly, that rationality might require probabilistic coherence that is approximate or better. My claim is still weaker: that rationality doesn't *require* radical probabilistic incoherence. This is also compatible with the possibility that rationality forbids *precise* probabilistic coherence (cf. Christensen 2007).

have had this view in mind.[329] That said, I do also want to argue that the view that belief is

credence 1 is not plausible, and that it does not provide a satisfactory way of defending the

deductive closure constraint.

Prima facie, the idea that belief is credence 1 view looks unpalatable, and the reasons

are familiar. We believe a lot of things, but it seems that we are maximally certain of very few

things. We both report ourselves as believing and outright assert a lot of things of which we

would not say that there is literally *no* chance that they are false. Moreover, it's easy to think of

situations that reveal our lack of maximal certainty in the things that we believe – for example,

bets at very bad odds.

One might think that someone sympathetic to the Williamsonian view of epistemic

probability mentioned in the section 5.3 – the view that knowledge has epistemic probability

1 – would also be sympathetic to the view that belief has credence 1. Williamson has at times

suggested that belief is something like knowledge "from the inside",[330] which might make us

think that if knowledge has epistemic probability 1, then belief has credence 1. Actually,

though, Williamson himself (at least in *Knowledge and its Limits*) seems to explicitly reject this

idea, on grounds basically similar to the standard grounds for rejecting the claim that belief

has credence 1:

> "We should question the association between evidential probability 1 and
> absolute certainty. For subjective Bayesians, probability 1 is the highest
> possible degree of belief, which presumably is absolute certainty. If one's
> credence in P is 1, one should be willing to accept a bet on which one gains a

---

[329] As noted below, Clarke (2013) is an exception. His view does have some partial forerunners. See Levi (1983: ch. 1) and (to a lesser extent) Van Fraassen (1995).

[330] See Williamson (2000: 44-47), where he tentatively proposes understanding believing p as treating p as if one knew p.

penny if P is true and is tortured horribly to death if P is false. Few propositions pass that test."[331]

However, Roger Clarke (2013) has recently defended a version of the view that belief is credence 1 which purports to get round these standard objections. His strategy for avoiding them involves appealing to what he calls "sensitivism" about degrees of belief, according to which a subject's degree of belief changes "from context to context, depending on the space of alternative possibilities." It is important to see here that although Clarke talks of "context", he does not (I think[332]) mean to endorse any *semantic* thesis about the word 'belief' (or 'credence') and its semantic value changing as the conversational context of the belief-*attributor* charges. Rather, the idea is that the subject's beliefs actually *change* depending on the salient "space of alternative possibilities" for *her* qua subject. This space of alternative possibilities may be affected by factors that are by the subject's own lights evidentially irrelevant, such as changes in what is practically at stake, or in what sort of reasoning the subject is engaged in.[333]

---

[331] Williamson (2000: 213). See also Williamson (2005: 685). Interestingly, this seems to commit Williamson to joining me in affirming at least that rationality does not require you to line your *credences* up with the evidential probabilities. Of course, if one thinks that credence and evidential probabilities must line up, one can just as well take this to be an argument against the view that knowledge has epistemic probability 1 as one can take it to be an argument for the view that belief is credence 1. So, for example, Greco (2013) argues that these standard objections to the view that belief is credence 1 generalize to the view that knowledge has epistemic probability 1.

[332] Confusingly, Clarke writes elsewhere in describing his view: "belief is context-sensitive…an agent may count as believing that p in one context but not another, without any change in her doxastic state between the two contexts" (Clarke forthcoming: 9). This makes it sound much more like Clarke has a semantic thesis in mind. But in that case I do not understand how Clarke could make the other claims that he does, like that when a subject believes p depends on whether *she* is taking any not-p possibilities seriously in the context. And in turn, this would prevent Clarke from being able to make his central claim that belief is credence one.

[333] Of course, everyone will agree that a subject's doxastic states can change as her situation changes, if this includes her (perceived) evidence changing; moreover, everyone will also agree that a subject's doxastic states can change without a change in her (perceived) evidence, such as when she is subject to various psychological biases and salience effects, as what is practically at stake changes (cf. Nagel 2008). This makes it surprisingly hard to say what the distinctive claim here is. One natural way of getting a grip on the thesis that *belief* is sensitive in the relevant way might be to say that there can be a change in whether a subject believes without a change in her underlying credence, but that obviously won't do for Clarke, since he thinks that the underlying

This idea of the sensitivity of doxastic states in general is not new with Clarke; it has been defended before concerning outright belief.[334] What is perhaps new with Clarke is the idea that one's *credence* is sensitive to one's situation in a similar way.[335] On the picture that others have argued for before, one's underlying credence, or level of confidence, stays constant across situations (barring apparent changes of evidence), and what changes is just whether that level of confidence suffices for full belief given the situation (either descriptively, if the view is that belief just is having credence above some contextually variable threshold, or normatively, if it is that belief requires more than this).[336] Clarke's view is more radical: he thinks that one has credence 1 – and hence believes – a proposition when one does not *take seriously* any alternatives in some particular situation – but whether one takes alternatives seriously depends on the stakes, what is salient, what kind of reasoning one is engaged in, and

---

credence changes too. But it is at least distinctive of Clarke's view that he thinks that a subject's change in credence can be *rational* even when, by the subject's own lights, nothing evidentially relevant has changed.

[334] See Weatherson (2005) and Ganson (2008). It has also been defended, even more famously, as a thesis about (the non-doxastic component of) knowledge, most prominently by Hawthorne (2004), Stanley (2005), Fantl & McGrath (2009), and Weatherson (2012). Hawthorne suggests that something like this could help with the preface paradox: "in those settings in which we intuitively think of putting long conjunctions to work – where some practical issue turns on the whole long conjunction – knowledge [of the individual conjuncts] is destroyed" (*ibid*.: 183). One could adapt Hawthorne's suggestion to concern belief rather than knowledge given the view defended by Weatherson (2005). But without Clarke's claim that belief is credence 1, it's not clear how this suggestion resists the argument I have given against deductive closure.

The same goes for the sensitivist view of belief advocated by Leitgeb (2014b). For all its formal sophistication, Leitgeb's view, as I understand it, only avoids my argument by being committed to the horrible consequence that at any one given moment in time, a subject cannot count as having more than a tiny handful of probabilistically independent beliefs. Leitgeb does not make this consequence explicit, and it is easy to miss since he orients us to always thinking about a doxastic agent's view of the probability space with respect to some fairly narrowly-defined set of questions at any one particular moment. But once we try to extend it to an agent's total doxastic state, the horrible consequence comes into clearer view. I think that Leitgeb, too, would do better to embrace Clarke's view that belief is credence 1.

Some of the criticisms that I make of Clarke below pertain to his radical sensitivism, and thus carry over to Weatherson and Leitgeb also. For a fuller critique of a radically sensitivist account of belief, and an articulation of a compelling alternative that maintains a link between belief and practical dispositions, see Ross & Schroeder (2014).

[335] Although Stanley (2005) hints, at times, at such a thesis about rational credence.

[336] Ganson (2008: esp. 455) is especially clear on this point.

so on. So as one's situation changes, one shifts from assigning a proposition credence 1 (and hence believing it) to assign it credence < 1 (and hence not believing it). Since this view identifies belief with credence 1, it blocks the premise of weak fallibilism in the argument I gave against deductive closure.

I think this view is very problematic, for at least four reasons. First, and most simply, our ordinary belief-attributing practices seem to allow us to self-attribute beliefs even when we are explicitly acknowledging a possibility of error *simultaneously and in the same context*. For example, it sounds completely fine to say something of the form "I believe that P, but I'm not sure" or "I believe that P, but I might be wrong" or even "I believe that P, but it might be that Q," where the proposition that Q is an incompatible alternative to the proposition that P.[337] It's of course open to Clarke to stipulate a notion of belief as credence one (in context), but if the account doesn't do justice to the thing that ordinary belief-attributing practices refer to, then it will be of limited interest in resolving the preface-type puzzle. It will still turn out that this puzzle arises for the thing we ordinarily mean by 'belief'.

Second: consider some case where the stakes change and I start taking some counter-possibility seriously. Take some ordinary proposition that I believe, like the proposition that I was born in the month of August. Suppose now that I'm offered some very unfavorable bet on this proposition – I win $1 if it is true and lose everything I own if it is false. On Clarke's

---

[337] It might be replied that sentences of the form 'I believe that P' are not really belief-reports, but are rather "hedged assertions" that P (cf., e.g., Murray 2014). It might then be claimed that the real way to express a belief that P is simply to assert that P, and that assertions of the form "P, but it might be that Q" (where Q is, again, an incompatible alternative to P) do not sound good. I am skeptical that sentences of the form 'I believe that P' are never actually belief-reports. Moreover, there are ways of making it clearer that one is reporting a belief rather than hedging an assertion, for which the conjunctions still sound good to me: for example, one can say: 'my own view is that P, but it might be that Q' or even 'I firmly believe that P, but of course it's possible that I'm wrong'. Finally, there is a good alternative explanation of the badness of 'P, but it might be that Q' that does not appeal to some incompatibility between belief and recognizing a chance of error: it is predicted by the standard contextualist semantics for modals that I defended in chapter 3 above. If asserting P is a proposal to add P to the modal base, then the modal claim 'it might be that Q' can be seen as quantifying over only the P-worlds, in which case it will come out false. See Worsnip (forthcoming-a) for more.

account, I began with credence 1, since I do not ordinarily take the possibility that I was not born in the month of August seriously – but when I am offered the bet, I start taking those possibilities seriously, my credence changes, and I lose my belief.

The problem with this, I think, is as follows. A credence is (or at least, carries a commitment to being equal to) one's assessment of how likely or probable a proposition is to be true, in a suitably evidence-relative (it is not objective chance that I have in mind here) sense of 'likely'. So, when one changes one's credence for some proposition, one is thereby committed to thinking *either* that the degree of probability of the proposition in question has changed *or* that one made a mistake in one's previous estimate of that probability. But in the case where I am offered a bet on the proposition that I was born in August, I may not think either of these things. It is very unnatural for me to think that, after I have been offered the bet, the probability or likelihood that I was born in August has changed. But nor does it seem like, after I am offered the bet, I need realize that I made a *mistake* in my original assessment of the probability that I was born in August.

Of course, there will surely be *some* cases where a change in stakes causes me to revisit my original assessment of the probability, to attend to some possibilities that I had been overlooking, and to adjust an unreasonably overconfident credence.[338] But this had better not be Clarke's general story about what happens for ordinary propositions that we believe. For then it looks like, on Clarke's view, our ordinary beliefs are all unreasonable. Since belief, for Clarke, just is credence 1, in order to say that our ordinary beliefs are often reasonable, he must say that credence 1 is often reasonable. So it cannot be that what happens, whenever

---

[338] Again, see Nagel (2008).

there is a shift of situation, is that we realize that we had been unreasonable in assigning the proposition credence 1.

The much more natural thing for me to say, after I am offered the bet, is that I still take the probability that I was born in August to be just the same as I always did. Normally, that probability suffices for me to rely on the proposition, ignoring for practical purposes the relevant counter-possibilities. But in special, unusual cases, like the one where I am offered the unfavorable bet, it will not suffice. This is just the standard decision-theoretic picture: credences remain stable across situations, providing a baseline from which to calculate the expected utility of different courses of action as the situation changes.[339]

By contrast, Clarke wants to say that as long as I am *treating* something as certain, and ignoring the relevant counter-possibilities, I actually count as having credence 1 in the circumstances. The problem with this, however – and this is my third objection – is that in order to even work out whether it is OK to treat a proposition as certain, and to ignore the relevant counter-possibilities, I need some baseline estimate of the probability of that proposition. For example, I know that ordinarily, I can just treat the proposition that I was born in August as certain, and I also know that in special circumstances, like that of the unfavorable bet, I cannot do so. But in order to be in a position to know either of those things, I need a baseline estimate of what the probability that I was born in August is. This baseline estimated probability *explains* why it is OK to treat the proposition as certain in ordinary circumstances (if the probability was just 0.3, for example, it wouldn't be OK to treat it as certain), as well as explaining why it is not OK to treat the proposition as certain in special

---

[339] See also Greco (2013: 101-2).

circumstances.[340] Without such a fixed baseline estimate, I would have no grip on why the changes in circumstances call for me to start and stop taking counter-possibilities seriously.

Then we must admit that, in addition to talking of whether I am *treating* a proposition as certain, and ignoring counter-possibilities, we must also make room for a (in one sense more fundamental) notion of my actual estimate of the probability of the proposition. But that latter notion, I think, just is the one that decision theorists and epistemologists mean to pick out in talking of credence. Of course, one could stipulate a meaning for 'credence' where it just tracks what one is *treating* as certain at any particular time, such that when I temporarily ignore the counter-possibilities to p, I have credence 1 for p.[341] But the important point is that there will still be another notion, that of one's estimate of the probability of p, and that notion can then just be substituted in for 'credence' in our argument against deductive constraints. So Clarke's view does not succeed in evading that argument.

Fourth, and finally, I find it hard to understand how Clarke's view applies to non-occurrent beliefs. In a classical preface case, Clarke says that when the author is at work on the body of the book, she believes the claims in the book, but that when she is writing the preface, she is attending to the possibility of error, and thus does not believe the claims in the book (hence avoiding inconsistency with her claim that she has made errors).[342] But what about times at which she is not actively at work on her book at all – when she is not attending to the

---

[340] Compare Weatherson's (2005) view, on which to believe a proposition is to have a credence such that *were* one to conditionalize on that proposition (which involves treating it as having probability 1), this would make no difference to one's actions in the circumstances.

[341] This may map onto what Wedgwood (2012b) calls "practical credence", as opposed to "theoretical credence". Wedgwood identifies full belief with being disposed to have practical credence 1 in normal circumstances. Unlike Clarke's view, this does not block our argument against deductive closure. We can just use theoretical, rather than practical, credence for the purposes of the argument.

[342] See Clarke (forthcoming).

claims in the book or the preface?[343] I take it we have lots of beliefs that we are not attending to at any particular point in time. Yet it is hard to see what Clarke should say about the author's beliefs in such states. It seems arbitrary to say either that she does not believe the claims in her book at such times, or to say that she does not believe the claim that her book contains errors at such times.

Perhaps Clarke could solve this problem by appealing to belief *fragmentation*.[344] Here, the idea is not just that the agent has different beliefs relative to different points in time, but that the agent has different subsystems of beliefs at single points in time. In other words, relative to some particular purpose, goal, or inquiry, an agent may believe p, while relative to some other purpose, goal, or inquiry, the agent may not believe p – all at a single point in time. In that case, Clarke could say that rational agents will never be inconsistent *within* subsystems, even if their different subsystems of beliefs are inconsistent when put together. I find this kind of move to be a slightly too convenient way of preventing intuitively rational agents from counting as inconsistent. Whenever an inconsistency in beliefs appears, of course one could in principle propose a way of fragmenting the agent's beliefs so that they don't formally count as inconsistent. Given the dubious psychological reality of actually different "fragments" of the mind, it is unclear that this approach really counts the agent in any robust sense as lacking

---

[343] The same problem applies to other potential appeals to the sensitivity of belief (without the thesis that belief is credence 1) as attempts to defuse the preface paradox. See also section 5.7 below.

[344] Thanks to Dan Greco for this suggestion. For proposals along these lines to deal with other problems regarding belief, see Elga & Rayo (ms.) and Greco (2014c). For an earlier version of this proposal specifically in the context of a discussion of deductive inconsistency, see Stalnaker (1984: ch. 5). However, Stalnaker ultimately sees fragmentation as involving irrationality, and Clarke could not agree to this if he wants to maintain that the agent in the preface case is rational, as he appears to. Stalnaker's view is that in preface cases responsible subjects do not outright believe the claims in their books. As I've already argued, this response is at best satisfactory only in the original preface cases, and does not solve the underlying problem.

inconsistent beliefs rather than just redescribing the inconsistent agent so as to officially avoid counting her as such.[345]

I can think of one other way of trying to resist weak fallibilism,[346] which is to claim that preface paradox cases are in effect rational dilemmas. On this view, although one rationally ought to have lots of beliefs which one assigns a non-one credence to, and although one rationally ought not to be probabilistically coherent, one also rationally ought to obey deductive closure. Unfortunately, that means that one cannot be rational. Note that this is *not* the same view as that mentioned in sections 5.1 and 5.3 above, namely that there is a conflict between deductive closure as a coherence requirement of rationality on one hand, and responding to one's evidence on the other. That view, I have suggested, is undermined by the way we have generated the challenge to deductive closure from within a coherentist framework. Here, by contrast, the idea would be that there is a conflict *between coherence requirements*.

I have already suggested some disadvantages of allowing rational dilemmas in section 1.6. It is also worth remembering again that preface paradox situations are not limited to special cases involving books and prefaces; they can arise for any large set of beliefs. So this approach does not just posit rational dilemmas in some unusual cases; it posits them all over the place. I think it is strongly desirable to avoid this result.

Indeed, this case illustrates a point I made against rational dilemmas in section 1.6, namely that allowing rational dilemmas blocks our ability to rule out rational requirements. The argument I am giving effectively suggests that belief-credence coherence is a more non-

---

[345] See Norby (2014) for a more general version of this criticism of fragmentation models.

[346] Other than the nuclear option of denying that one can permissibly have a non-trivial number of beliefs that are non-trivially independent. See fns. 327 and 334.

negotiable requirement of rationality than deductive closure is, and notes that (given weak fallibilism) they are not co-tenable. If we reject weak fallibilism by allowing rational dilemmas, they become co-tenable – as do any two potential rational requirements! So one can never argue against one rational requirement from the greater plausibility of another, conflicting requirement. This makes it very hard to rule out any putative requirement of rationality, and makes positing such putative requirements effectively costless.

There is a danger, then, that we elevate features that it would be *nice* if agents' beliefs could have into genuine *requirements* of rationality. No doubt in one sense life would be easier if the deductive closure permission held, and we could always infer the logical entailments of our existing beliefs without any need to worry. No doubt in one sense things would be neater if our beliefs were deductively closed; their not being so is, in one way, regrettable. But this does not mean that we are irrational if we, noticing that we are not in an ideal world, knowingly violate the deductive closure constraint. When we lose the distinction between nice features of beliefs and rational requirements, we lose the ability to say anything about what agents rationally ought to do when the nice features conflict with the genuine rational requirements. Such cases are just written off as cases where the agent cannot do what rationality requires, so there is nothing to be said about what the agent rationally ought to do. We can do better than this.

5.6 Defending belief-credence coherence

One might wonder why the credence mentioned in belief-credence coherence is 0.5 in particular. The answer is that at the point that one has a credence lower than 0.5 in a proposition, one judges it to be more likely to be false than true. This attitude, I claim, is rationally incompatible with belief. Part of what it is to believe something is to take it to be

true – where the relevant contrast is it's being false. It is incoherent both to take something to be true and to judge it to be more likely to be false than true.

This leaves open the possibility that there could be more demanding rational constraints on the relationship between belief and credence that are more demanding than belief-credence coherence. For example, perhaps it is also irrational to believe something and to have a credence of exactly 0.5 in it, since the latter attitude still involves taking it to be no more likely to be true than false. I actually think that this is right – but I need not commit to it here. All I claim here is that having a credence lower than 0.5 in p and believing p is an incoherent combination. Moreover, if someone wants to protest that the relevant level of credence that should feature in belief-credence coherence is in fact *lower* than 0.5, the argument can still ultimately be run. However, I think that 0.5 is a good, non-arbitrary point to fix on, because it is the point at which one changes one judgment about whether p or not-p is more likely.

The incoherence involved in violating belief-credence coherence can be brought out, as a few different philosophers have noted, by considering the Moore-paradoxical flavor of utterances that report the two mental states jointly.[347] It seems incoherent to say "P, but it's more likely that not-P" or to say "I believe that P, but it's more likely that not-P".[348] However, Mark Kaplan (1996) has defended a view on which one can (rationally) believe a proposition p even when one has a credence lower than 0.5 in p, even though he acknowledges the Moore-

---

[347] See Christensen (2004: 48-49); Kaplan (1996: 142-143). This Moore-paradoxicality is closely related to that which I noted in defending another rational requirement, (ILC), in section 1.3 above. In fact, it is possible that one could get the case against deductive closure using (ILC) instead of belief-credence coherence. Still, I prefer the case that uses belief-credence coherence, since it seems easier for agents to systematically lack higher-order beliefs about their evidence for their first-order beliefs than to lack credences.

[348] See fn. 307 above on my notation here.

paradoxicality of such reports. He claims that such Moore-paradoxicality results from a confusion in the ordinary concept of belief.

Kaplan offers an interesting argument by analogy that there is really no incoherence involved in such a combination of mental states.[349] The analogy is a bet where one is offered either a certain $5, or a bet whereby one wins $10 if p is true but loses $1000 if p is false. Suppose that one is 0.95 confident that p. Then the expected value of the bet is lower than the certain $5. In that case, one might pick the certain $5. Nevertheless, one is 0.95 confident that taking the bet would maximize one's actual winnings. Kaplan points out, rightly, it seems to me, that there is no incoherence here. He then claims that, in preface-like cases, combining a belief in p (where p is, for example, a long conjunction of propositions) with high confidence that p is false is just the same.

This, I do not find convincing. There are two major points of disanalogy. First, to believe p is to believe that p is true. That much I take to be a non-negotiable aspect of belief, however much Kaplan charges the ordinary concept of belief with incoherence.[350] On the other hand, to choose an option is clearly not to believe that this option will have the best actual outcome of all the available options. Precisely because one makes decisions under risk, we are perfectly familiar with the possibility of hedging our bets and choosing something that we think probably won't end up having the best outcome of the available options, but protects us from a small chance of disaster. That is exactly what happens in Kaplan's simple betting case. Since choosing the option does not necessarily involve thinking it will have the best possible outcome, it does not clash with thinking that some other incompatible option is more

---

[349] Kaplan (1996: 144-145).

[350] See Section 4.5.

likely to have the best possible outcome. But since believing something does involve believing it to be true, it does clash with thinking that some incompatible proposition (such as its negation) is more likely to be true.

Second, and relatedly, in the case of the bet, even though one cannot say that declining the bet is more likely to have the best actual outcome of the available options, one can say that declining the bet has the best expected outcome of the available options.[351] Yet there is nothing analogous that one can say in the case of the belief that one judges to be more likely to be false than true. It's not as though one can say that, even though it is more likely to actually turn out to be false than true, the "expected truth-value" of the proposition is still *true*. Nor is belief in p somehow a good way of "hedging one's bets" as it is in the betting case. If anything were such a hedge, it would be choosing to suspend judgment about p even though one judges it very likely that p is false; it would not be to believe p even though one judges it very likely that p is false.

Funnily enough, Kaplan's own view of belief does not, on the face of it, seem hospitable to belief in things one judges to be more likely to be false than true. On Kaplan's view of belief, which he labels the "assertion view", "you count as believing p just if, were your sole aim to assert the truth (as it pertains to p), and your only options were to assert that p, assert that not-p or make neither assertion, you would prefer to assert that p."[352] One would think that, in any case whereby one judges that p is more likely to be false than true, to the

---

[351] In the sense of "expected" used in "expected utility" theory, where both the probability of a potential outcome and its goodness (or badness) are both taken into account. As simple examples like Kaplan's show, taking an option to have greater expected utility than another option does not necessarily involve "expecting" that it will have a better actual outcome. The latter is determined simply by the probability that it will have a better actual outcome, whereas the former also takes into account *how* good the better outcome would be as compared with the worse outcome.

[352] Kaplan (1996: 109).

extent that one's sole aim is to assert the truth as it pertains to p, one would (at least, if one is rational) *not* prefer to assert p!

As far as I can tell, Kaplan thinks that this is somehow mitigated by the fact that one is aiming to tell a comprehensive story about the world.[353] The thought seems to be this: there are lots of propositions, each of which I would prefer to assert than to assert their negations or to make neither assertion. So, given that I aim to tell a comprehensive story about the world, plus the assertion view of belief, I will count as believing many of these things. So far, so good.

The problem, however, for Kaplan, is to get from this claim to the claim that I believe, or would be rational to believe, the logical consequences of these individual claims (for example, a long conjunction of all of them). It seems to me that Kaplan must be moving from the claim that I would be willing to assert each proposition individually to the claim that I would be willing to assert their conjunction. (Otherwise, I cannot see how he would derive the claim that I believe the conjunction, according to the assertion account of belief, despite taking that proposition to be more likely to be false than true.) But that just doesn't follow.

Indeed, this move will seem unpersuasive especially to someone skeptical about the deductive closure constraint, who thinks there is a big, and important, difference between believing a set of individual propositions and believing their conjunction (see section 5.7 below). For why should the same difference not also be plausible with respect to assertion? If anything, in fact, the point is more obvious with respect to assertion: while it may be hard to put one's finger on what it is to believe a set of individual propositions, but not believe their

---

[353] See Kaplan (1996: 111); see also Christensen (2004: 71-73), who interprets Kaplan as holding this and, in effect, giving up the assertion account of belief as he has stated it. Christensen offers a nice critique of the alternative view that emerges; here, however, I focus on diagnosing why I think Kaplan mistakenly thinks that his ultimate position can be squared with the assertion account as he has stated it.

conjunction, it is *easy* to put one's finger on what it is to be disposed to assert any one of a set of individual propositions, but not disposed to assert their conjunction.[354] And it is especially easy to see why one might be in that position if one thinks about a person whose assertions are governed by the sole aim of asserting the truth as it pertains to the proposition in question. For, as the preface paradox cases teach us, it could be that with respect to each individual proposition, I am very likely to be asserting the truth, but with respect to the conjunctive proposition, I am not. In this way, we see that the assertion account of belief actually *strengthens* the case against deductive closure.

5.7 Living without closure in the theory of belief

This completes my positive case against deductive closure. As I stressed at the start, this case is generated internally from within a coherentist theory of rationality. This undermines the possible strategy of claiming that deductive closure is a coherence requirement, albeit one that sometimes conflicts with our evidential reasons for belief. Moreover, the challenge to deductive closure that I have developed is especially robust, because it does not assume any particular theory about the reducibility of belief to credence, or any particular theory of evidential support.

The challenge also calls into question whether it is even *permissible* to have deductively closed beliefs. Recall that in section 5.2 I distinguish the deductive closure constraint from the deductive closure permission. The challenge as developed here suggests that there are cases where one can have a body of beliefs that are not irrational, where believing some logical

---

[354] This parallels the analogous point in the next section about what one is disposed to rely on in one's reasoning.

consequence of those beliefs does make one irrational. That means that it falsifies the deductive closure permission as stated in section 5.2.

In section 5.8, I will turn to the question of whether deductive *consistency* can nevertheless be salvaged as a requirement of rationality. Before that, I want to deal with one issue that arises with the rejection of deductive closure that might lead to some residual reluctance to reject deductive closure (and in particular the deductive closure permission). One might think that it is at least partly constitutive of believing p that one is disposed to use p in one's reasoning, including one's theoretical reasoning.[355] And so, one might think that, if one really does (for example) both believe p and believe q, then it has to be the case that one would be disposed to come to believe their conjunction at least if one considered it. This might make it hard to see how it could be permissible to believe p and believe q without it being permissible to believe their conjunction (and similarly for more complex logical entailments).[356] Relatedly, it may seem to bring no real new commitment to believe their conjunction, as distinct from believing both of them individually – after all, the conjunction is only as logically strong as its conjuncts.

My answer is that one can be disposed to use some proposition in one's reasoning, and disposed to use some other proposition q in one's reasoning, without being disposed to use them jointly. Here is a case to illustrate this kind of possibility. Consider the experience of listening to philosophy colloquium talks. Compare the talks of two speakers, Amanda and Barbara. Both argue for (different) philosophical theses about which you, coming into the talks, have no settled opinion about. Suppose, further, in a rather idealized manner, that there

---

[355] For various contrasting developments of this idea, see amongst many others, Stalnaker (1984); Williamson (2000: 99); Schwitzgebel (2002); Weatherson (2005), Fantl & McGrath (2009: ch. 5); Ross & Schroeder (2014).

[356] Thanks to both Tamar Gendler and Bruno Whittle for posing this objection to me in slightly different forms. Foley (1993: 166-8) also voices a worry somewhere in the neighborhood of this one.

is a list of (say) a few hundred philosophical claims that you are confident of and that, for each, you would comfortably describe yourself as believing the claim in question if asked. We'll call it "your list", but you need not be keeping active track of the list. In her talk, Amanda shows that her thesis follows from just three premises, all of which are on your list. Barbara, meanwhile, shows that her thesis follows from 150 premises, all of which are on your list.

I think it is not at all difficult to imagine the case such that you will be disposed to come to believe Amanda's thesis but not Barbara's. But the problem need not be that Barbara's individual premises are *individually* less secure than Amanda's. It's just that there are so many of them. We can imagine the case such that, for any three premises out of the 150 that Barbara uses, you would be disposed to come to believe something that Amanda (or Barbara) showed to follow from those three premises. However, when Barbara relies on all 150, you are less apt to be persuaded. "So much has to break the right way for her argument to succeed," you might think, "so I'm going to continue to suspend judgment about her conclusion."

Clearly there are structural analogies between this case and the preface. But this case also brings out how one can be disposed to rely on a number of propositions *individually* in one's reasoning without being disposed to rely on them *jointly*. So, we can concede that there may be a constitutive tie between belief and what one is disposed to rely on in one's reasoning, without conceding that this means one must be disposed to rely on one's beliefs jointly in one's reasoning. If that's right, there is no puzzle about how one could have individual beliefs without being disposed to come to believe their logical consequences, and thus about how the deductive closure permission could be false.

Of course, it is also important that we are often disposed to put our beliefs together in our reasoning. I propose that we can capture this by saying that when one is disposed to put two beliefs together in one's reasoning, one is rightly represented as (at least implicitly)

believing their conjunction. In other words, the test of whether a subject believes (p & q), rather than just believing p and believing q, is whether the subject is disposed not just to use p and q in her reasoning individually, but to combine them.

This does have one result that initially seems at least slightly counterintuitive. In order to perform a conjunction operation, and move from belief in p and belief in q to believing (p & q), one must already be disposed to rely on p and q jointly in one's reasoning. But according to the present account, if one is already disposed to rely on p and q jointly in one's reasoning, one already counts as believing the conjunction. So it seems that one can never come to believe a conjunction by means of performing a conjunction operation.[357] The solution, I think, is to maintain that performing conjunction operations, in cases where you are disposed to, is a process of making explicit what you already implicitly believe. This doesn't mean that by believing p and believing q, one already implicitly believes (p & q).[358] On the contrary: I have been trying to argue that one can have those individual beliefs without even being disposed to infer (p & q). Rather, the claim is that *when one is* in a state where one is disposed to rely on p and q jointly to infer (p & q), one already implicitly believes (p & q).[359]

In many cases, we just are disposed to jointly rely on our beliefs, without much worry about the aggregation of risk. That is as it should be: working out the exact probabilities given our individual credences would often be a pedantic waste of time. But we'd better at least have an over-ride such that when risk-aggregation does really become problematic, we won't rely

---

[357] Thanks to Mark Maxwell for helping me to pinpoint this objection.

[358] Some views of belief do seem to have that consequence: in particular the "possible worlds" picture advocated by Hintikka (1962) and Stalnaker (1984). I examine the relation between these views and preface cases, arguing that the former cannot stand up to the pressure generated by the latter, in Worsnip (ms.).

[359] As with similar cases of implicit belief of propositions you have never considered, but would be disposed to explicitly believe upon consideration – for example, the proposition that the Amazon river is more than 56 feet in length.

on a lot of propositions jointly. And it seems that we do have such an over-ride, since we aren't always disposed to jointly rely on such propositions, as shown by the example of Barbara's philosophy talk. Or consider another case, where it is practical reasoning that is at stake. Consider a man who is having an affair. On some one-off occasion where he has an opportunity to meet his lover at home, he might just rely in his reasoning on the proposition that his wife won't suddenly and unexpected come home from work today. But he'd be crazy to plan to do this every day by conjoining his belief with the proposition that his wife won't be coming home tomorrow, or the day after, or the day after that, and so on. This is so even if these propositions were to individually enjoy credences greater than or equal to his credence for her not coming home today – such that for any future date, if the one-off was something he was planning for that date, he'd rely on the proposition that she won't come home that day. The risk of error aggregates; if he does this every day, it's overwhelmingly likely that eventually he'll get caught.

Note how this picture contrasts with one on which, as soon as one ceases to rely on a proposition even jointly with other propositions, one ceases to count as believing it entirely.[360] That results in an exceptionally unstable view of belief, on which as soon as one starts to consider long conjunctions of one's beliefs, those beliefs themselves suddenly disappear. Nevertheless, it might be thought that such an account is necessary, if we are inclined to tie belief to the dispositions involved in reasoning. This is too quick, however.[361] We can instead say that one continues to count as believing a proposition as long as one is *disposed* to rely on it individually, even when one is unwilling to rely on it jointly with other propositions.

---

[360] Cf., e.g., Weatherson (2005).

[361] Here I agree with, and build on, the criticisms of this view developed by Ross & Schroeder (2014). My alternative resembles theirs in a number of respects.

As the cases of the philosophy talks and of the cheating husband show, one can easily be simultaneously be disposed to rely on several propositions individually, while being disposed not to rely on them jointly. In fact, we can even cook up a story where both dispositions are simultaneously activated.[362] Suppose that in Barbara's talk, she makes an argument for a conclusion C, making use of 150 premises, and in doing so also proves that some intermediate conclusion IC follows from just 3 of the premises. In coming to believe IC but not C, as one very well might, one's disposition to rely on the propositions individually (and jointly in small numbers) and one's disposition not to rely on them jointly in large numbers are both activated. The present view explains this by saying that you believe the individual propositions (and short conjunctions of them), but not the long conjunction of all the propositions. That is a clean explanation that does not require your beliefs to fluctuate continuously as the subject of your reasoning shifts during the talk, or to be continuously fragmented in some way. Since the state you are in does not seem conflicted or turbulent, I think this is an advantage. In any case, it shows that the view on which your beliefs are in constant flux is far from the only way of making good on the idea that there is a tie between belief and dispositions to reason and act.

5.8 Consistency without closure?

Again, the target of the attack so far has been deductive closure. One might wonder if one can still defend a coherence requirement of deductive *consistency*, which is weaker than a closure requirement. The putative requirement of logical consistency can be formulated as follows:

---

[362] For a related attempt to exploit the simultaneity of two different dispositions being activated to put pressure on a sensitivist account, see Greco (2013: 99-100).

Deductive consistency constraint. Rationality requires of S that if S's total belief set logically entails some proposition p, then S does not believe not-p.

Since I have argued that having deductively closed beliefs is sometimes not even *permissible*, the idea here would have to be that in at least some preface-type cases we are required to suspend judgment about the logical entailments of our beliefs: to neither believe nor withhold.

The deductive consistency constraint cannot, I think, be argued against in a way exactly analogous to my argument against the deductive closure constraint. In that argument, I appealed to a requirement forbidding agents from believing propositions that they assign credence below a minimum threshold of 0.5. One might hope there would be a parallel requirement requiring agents to believe those propositions that they assign a credence *above* some maximum threshold. In fact, I think there is such a requirement, but the threshold, I think, is 1:

Belief-credence coherence (max). Rationality requires of S that if S has credence 1 in p, then S believes p.[363]

Belief-credence coherence (max) specifies a maximum lower bound at which having the credence in question (1) but withholding belief is irrational. Unfortunately, this requirement is of no use in arguing against deductive consistency. A probabilistically coherent agent will have a very low credence in the conjunction of a set of beliefs each of which she has high, but non-1 credence in. But the credence will not be 0, and so her credence for its negation will not be

---

[363] Obviously, the requirement does not say that one may *only* believe those things for which one has credence 1 – I argued against that view in section 5.5.

1. So one cannot generate the result that rationality requires her to be deductively inconsistent in this manner.

Nevertheless, I still think that in light of the failure of the deductive closure constraint and in particular the deductive closure permission, the deductive consistency constraint loses plausibility. Suppose that one finds oneself with a set of beliefs, each of which one has a very high credence for (without irrationality), but which one discovers to be inconsistent. Given the deductive consistency constraint, one must suspend judgment on at least one of these propositions. But which one? Picking a belief from the set at random seems arbitrary. Notice that the answer here cannot just be: you should eliminate the belief that is individually irrational, even though it's not transparent to you which it is. For the case is not one where you think that one of the beliefs is likely *irrational*. Rather, it is a case where you simply think that one of the beliefs is *false*. But we can stipulate that you have no good grounds for figuring out which of the beliefs is false: each individually is highly likely.

This point is intensified if we think – as suggested in the previous section – that belief is tied to what one is willing to rely on in one's reasoning. Picking a proposition at random to suspend judgment on is not just a matter of taking it off some official belief list and going on as usual. It involves ceasing to use that proposition in one's reasoning as one does one's other beliefs. To pick an arbitrary belief to do this with seems bizarre. What seems much better is to continue to rely on each proposition individually, but being wary of putting them together – in much the way sketched in the previous section.

These considerations are not as decisive, in my view, as the ones adduced against the deductive closure constraint in this chapter. Still, they are in my view enough not to impose deductive consistency as a coherence constraint of rationality either. Part of the point of this dissertation has been to show that there are a number of coherence constraints involving

outright belief that are not deductive constraints. In this chapter so far, we have encountered some of these, ones that relate outright belief to credences in certain ways. In the final section, I will consider whether having rejected both deductive closure and deductive consistency, there is any room for accepting other deductive or quasi-deductive constraints.

5.9 Other deductive constraints?

So far in this chapter I have argued against deductive constraints of closure and consistency. However, it is worth considering whether there are weaker deductive constraints that might still be acceptable.

There is at least one such constraint: a non-contradiction requirement:

Noncontradiction requirement. Rationality requires of S that if S believes p, then S does not believe not-p.

Whereas the deductive consistency constraint forbade large inconsistent sets of propositions, the non-contradiction requirement forbids believing two propositions that strictly contradict each other – a proposition and its negation.

The noncontradiction requirement is not vulnerable to the same kinds of worries as the deductive closure and consistency constraints, because it is plausible that one cannot rationally assign both a proposition and its negation a credence of greater than 0.5. Consequently, the case in which one believes contradictory propositions could not be one in which one has high credences in both propositions without irrationality, while taking it that they cannot both be true. Indeed, if one accepts a slightly strengthened version of belief-

credence coherence, on which it is irrational to believe a proposition that one does not assign a credence greater than 0.5 to, and one accepts that one cannot rationally assign both a proposition and its negation a credence of greater than 0.5, then it just follows that one cannot rationally believe both a proposition and its negation. Moreover, there is a clear and intuitive incoherence to saying that one believes both a proposition and its negation. I am happy to accept the noncontradiction requirement.

What about inconsistencies or closure failures involving other small sets of beliefs? For example, suppose one says that one believes p, and that one believes (p→q), but that one does not believe q. Hasn't something gone wrong here? Isn't there at least *usually* an incoherence or failure of intelligibility in this sort of case?

I share the sense that something has gone wrong here, but unfortunately it is unclear how to capture this sense without overcommitting to strong closure or consistency requirements of the sort already rejected in this chapter. As several authors have argued,[364] restricting the relevant requirement to the simplest or most obvious logical entailments may not help. Even in the simplest case, where one single proposition entails another, one can construct a chain of deductions, where the risk that one has made some reasoning error aggregates as one performs more and more single-premise deductions. Moreover, conjunction is itself a very simple logical operation, moving just from believing p and believing q to believing (p & q). But it leads us into the preface paradox in the most notorious way.

One might try to insert *ceteris paribus* conditions into the relevant requirement. So we could say, for example, that rationality requires that if S's total belief set entails p, and S does not have any excusing explanation of why she does not believe p (such as one involving

---

[364] See Lasonen-Aarnio (2008); Schechter (2013); DeRose (forthcoming: ch. 11).

considerations of risk aggregation), and is actively considering whether p, then S believes p. We might weaken this yet further, so that the antecedent reads "if S's *believes* that S's total belief set entails p…" This is a possible way of going, although I do not think we should be sanguine about filling out the details. One disadvantage of this way of going is that it makes references to the notion of an excusing explanation, a notion which calls out for explanation. Moreover, if the right way to fill that out turns on a thickly normative notion of what really counts as a good excuse and what doesn't, it's not clear that the requirement will count as a genuine *coherence* requirement. It will make whether an agent violates the requirement depend on whether she has a genuinely good excuse for not believing q, and then it is arguably not simply about the relations between her attitudinal mental states. One could make the requirement refer to an agent simply *believing* she has an excuse for not believing q. That would restore its status as a coherence requirement. But some may find it too forgiving to adequately capture the phenomenon on its own.

I remain open to the possibility that there is a way of getting this requirement right. Unfortunately, I do not have space to settle it here. It is an important question for future work.

Conclusion

In this dissertation, I have tried to show that coherence requirements, as distinct from evidential reasons, do significant work in epistemology, in debates about higher-order evidence, doxastic voluntarism and deductive constraints, amongst others. In this conclusion, I want to look forward to some other areas of epistemology where I think coherence requirements can help us to make progress, and explain (in a somewhat more sketch-like fashion) how they do so.

(a) Conditionalization

In the debate over rules for belief revision, the status and content of the correct rule for Bayesian "conditionalization" is a matter of dispute. Conditionalization rules tell you how to update your credences when you gain new information. Specifically, they tell you to update on your conditional "priors". Your conditional priors give your credences for propositions *conditional* on the truth of other propositions. For example, you might have a prior credence for p conditional on q of 0.7. The conditionalization norm says, very roughly, that if you then learn q (and nothing more), then your new unconditional credence for p should be 0.7.

In the debate about Bayesian conditionalization, there is a polarization between two camps – "subjective" and "objective" Bayesians. In crude terms, subjective Bayesians hold that (close to) anything goes when it comes to one's priors. Provided that one obeys constraints of probabilistic coherence, the idea is, there are no constraints on what one's priors

should be. But whatever one's priors actually are, one should update on those actual priors.[365]
In equally crude terms, objective Bayesians (of the most extreme kind) hold that one should
not update on one's *actual* priors, but rather on the objectively correct set of priors, which
specify what credence to have given any potential set of evidence.[366]

On the objective view, so-called "conditionalization" turns out to amount to no more
or less than a requirement to believe what one's evidence supports – our old friend (ER). This
entirely loses the idea that there is something distinctively rational about correctly
conditionalizing on one's own priors. If one's own priors do not match the objectively correct
set of priors, on this view, one has no hope of conditionalizing correctly; there is no rational
requirement that one complies with when one conditionalizes on objectively incorrect priors.

Less obviously, some subjective Bayesians take themselves to be giving a theory of
how to believe what the evidence supports.[367] Here the idea is that evidential support is not
actually a two-place relation between a proposition (or body of propositions) and a doxastic
attitude, but rather a three-place relation that also involves the believing subject's standards
(or prior probability function). So, what it is for a proposition p to evidentially support having
a high credence in a proposition q *for a subject* is just for that subject's conditional prior
probability for q conditional on p to be high. This, it must be said, is a *radically* error-theoretic
view of evidential support, relative to our ordinary judgments. For example, taken again the
example discussed in chapters 2 and 4 of the fairies in the garden. Suppose that I have a high
subjective prior for the proposition that there is a magical garden fairy, conditional on there
being dew on the grass. The present view suggests that, when it becomes part of my evidence

[365] Cf., e.g., Jeffrey (2004); Joyce (2005); Titelbaum (2010).

[366] Cf., e.g., T. Williamson (2000: ch. 10); J. Williamson (2010); Hedden (forthcoming).

[367] Cf., e.g., Joyce (2005); Titelbaum (2010).

that there is dew on the grass, my evidence now supports having a very high credence that there is a magical garden fairy. This is a very radical kind of subjectivism about evidential support, one that is highly revisionary of our ordinary intuitions. We normally assume that subjects are capable of being wrong about what their evidence supports (and not merely because they don't know what their own subjective priors are). In this case, we'd normally say that I am simply wrong that the dew on the grass supports having a credence that there is a magical garden fairy.

The option of construing conditionalization as a wide-scope coherence requirement allows us to avoid both of these extremes, in much the same way that wide-scoping about other requirements allowed us to avoid other false dichotomies (cf., e.g., section 0.3). On this view, we give up the harmful assumption shared by both camps that conditionalization is a theory of how to respond to one's evidence, and simply construe it more modestly as a principle of coherence between one's prior, conditional credences, one's beliefs about what one has learnt, and one's unconditional credences. On this view, conditionalization is (roughly) as follows:

> **Wide-scope conditionalization, very roughly.** Rationality requires that if one has a prior credence C for p conditional on q, and one believes that one has learnt q (and q only), then one has an unconditional credence C for p.

The requirement refers to one's *actual* priors, and not to some objectively correct set of priors. So, unlike the "objective" Bayesian view, it preserves the thought that when one updates on one's own priors, one has successfully complied with a requirement of rationality. Equally, however, unlike the "subjective" Bayesian view, it is wide-scope, and so it doesn't let one's

306

priors simply *dictate* what it is rational to believe. One cannot, just by having crazy priors, make it the case that rationality requires one to have correspondingly crazy unconditional credences. Still less can one make it the case that one's evidence supports having such unconditional credences.

I hope that this gives some illustration of how the wide-scope coherence account could help here, and again, how epistemology has been harmed by failing to distinguish coherence requirements from principles about evidence-responsiveness. However, there is a lot more to be said about how to work out the account of conditionalization as a wide-scope coherence requirement. In particular, a very important question concerns whether to construe conditionalization as a synchronic or a diachronic coherence requirement. In section 3.7, part b, I argued that the sorts of requirements that occupied us in this dissertation should be construed synchronically, since the only plausible diachronic versions of such requirements turn out to add nothing new to their synchronic versions. However, one may think that conditionalization is a different case. If conditionalization is a diachronic requirement, then, as I argued in section 3.8, we may get the result that once one's prior is fixed at some initial point in time, the only way to satisfy the rational requirement is to update on that prior at a later point in time; the original prior is now unrevisable. And then we actually do get the result that one is rationally required to update on one's priors, no matter how crazy those priors.

If this seems implausible, perhaps we should explore construing conditionalization synchronically – without letting it collapse into a requirement to have the credences that one's evidence supports.[368] Perhaps we can understand one's priors in some non-temporal sense, such that one's priors are revisable, even as one considers updating on them. Then

---

[368] As Hedden's (forthcoming) version of "synchronic conditionalization" does.

conditionalization merely forbids *simultaneous* mismatch between one's priors and one's unconditional credences.[369] I do not want to take a stand on this here. There is much more to be said.

(b) Peer disagreement

For another place where a wide-scope, coherentist view can help us make progress, consider the recently "hot" epistemological debate on how one ought to respond to peer disagreement – that is, disagreements with those that are (or appear to be) as intelligent, intellectually virtuous, and otherwise reliable as oneself.

In this debate, there is a dispute between so-called "conciliationist" and "right reasons" views. On the conciliationist view, when one encounters peer disagreement, where there is a gap between one's own credence in some proposition and that of one's disputant, one is always (or at least typically) required to revise one's credence to some level intermediate between one's original credence and that of one's disputant.[370] On the most extreme version of this view, the "equal weight" view, one's new credence should be equidistant between one's original credence and that of one's disputant.[371]

On the "right reasons" view, by contrast, what one should do simply depends on whether one's original credence represented a correct response to one's evidence. If it did, one need not make much revision to one's credence; if it did not, one must revise one's credence.

---

[369] But in turn, if conditionalization is construed synchronically, our theory may have the result that rational agents can still be vulnerable to Dutch books. This is no doubt a cost.

[370] See, e.g., Christensen (2011); Elga (2007); Feldman (2006).

[371] Cf. Elga (2007).

On the most extreme version of this view, the disagreement itself provides *no* reason to change one's credence: at most it could provide reason to reconsider whether one's credence was the right response to the evidence. Any need to actually revise one's credence is down to an originally unreasonable response to one's evidence.[372] On a less extreme version of the view (sometimes called the "total evidence" view), the disagreement itself is *some* evidence against one's original view, but it might be relatively insignificant relative to one's original evidence, and there is certainly no general formula for saying where one's credence should end up relative to one's original credence and that of one's disputant.[373]

One criticism that theorists in the right reasons/total evidence camp make of those in the conciliationist camp is voiced by Kelly (2010: 122-8).[374] Kelly asks us to imagine a case in which you and I share the same evidence about whether p, yet I arrive at a credence of 0.7 and you arrive at a credence of 0.9. We then learn of our disagreement. According to the equal weight view, Kelly notes, we should then both revise our respective credences to 0.8. But then it looks like our original evidence simply has no bearing on what credence either of us ought to have. For notice that in the set-up of this case, we said nothing about what the evidence supported. Without any sense of what the evidence supported, the equal weight view seems to have given us a verdict about what credence we ought to have. Surely, Kelly says, this can't be right. Indeed, he suggests, the point is sharpened if we suppose that *both* of us badly mistook our evidence; in fact, it supports a much lower credence than either 0.7 or 0.9. But the equal weight view, Kelly claims, says that we now each ought to have credence 0.8.

---

[372] See, e.g., Kelly (2005); Titelbaum (2015b).

[373] See, e.g., Kelly (2010); Lackey (2010); Weatherson (ms.).

[374] A similar objection is voiced by Weatherson (ms.).

However, this criticism depends upon interpreting the equal weight principle so that it is (i) narrow-scope[375] and (ii) uses a concept broad enough to exclude any other kind of normative evaluation. Kelly is assuming that the principle says that if you and I begin with credences of 0.7 and 0.9, and we are (or regard each other as?) peers, then we ought to each have a credence of 0.8. Suppose, however, that we read the equal weight principle as a wide-scope coherence requirement; very roughly:

Wide-scope equal weight principle, very roughly. Rationality requires that if one has credence C1 for proposition p, and one believes that another doxastic agent A is one's peer, and one learns that A has credence C2 for proposition p, then one revises one's credence to ½(C1+C2).

Suppose that one's original credence C1 is a highly unreasonable response to one's evidence. The wide-scoped equal weight principle does not commit us to the result that the credence you will end up with by conforming with this requirement is all-things-considered reasonable. All it commits us to is that you will have conformed with this particular rational requirement. As in the case of conditionalization above, this construal of the equal weight principle is modest in that it does not attempt to yield results about what is all-things-considered reasonable to believe on its own. This is for two reasons – because it is wide-scope, and so does not allow for simple forms of "detaching", and because it employs a single normative notion – that of a rational requirement – that need not be thought of as the only kind of normative evaluation available. So, *contra* Kelly, the requirement does not cause evidence to

---

[375] Kelly even calls the problem one of "bootstrapping" (Kelly 2010: 125); it is just analogous to the bootstrapping objections to narrow-scope versions of rational requirements in section 3.2 above.

drop out of the picture. Indeed, on an attractive way of thinking about things, the evidence has a major role to play in determining one's original, starting credence. The equal weight principle is restricted to saying how this credence should be *revised* once one encounters disagreement. Put somewhat loosely: if you feed crazy inputs into a rational process of belief revision, you may get crazy outputs; that is the fault of the inputs, not the process.[376]

Notice that the wide-scope version of the equal weight principle, as I have stated it, makes another crucial revision to a narrow-scoped analogue. The wide-scoped version of the principle allows that one could satisfy the requirement by revising one's belief that one's disputant is one's peer,[377] rather than revising one's credence in the proposition under dispute. According to most proponents of conciliationism, this is (at least typically) *not* a permissible response to disagreement.[378] If one is moved by this thought, one could take the belief in peerhood outside the scope of the 'requires' operator, leaving us with:

---

[376] Cf. Christensen (2011: 3-5) on this general point, though Christensen does not identify wide-scoping as a framework for explaining why the equal weight view avoids these consequences. Kelly considers the idea that conforming with the equal weight principle is not sufficient for one's ultimate credence being all-things-considered reasonable. To this he objects as follows. Suppose that you and I begin with credence 0.2 and 0.8 respectively. Suppose that your original credence is a correct response to your evidence, but mine is not. Then, if we conform to the equal weight principle, each of us will end up with credence 0.5. Since you began with a reasonable credence, your ultimate credence is also reasonable; mine, on the other hand, isn't. But, Kelly asks, if we have the same evidence and the same credence, how could this be right? I do not see why this is supposed to be a problem. There are many ways that one can end up at doxastic states that *would* be those of an epistemically ideal agent without being ideal oneself: for example, one could just quite *accidentally* get things right, as Kelly himself stresses elsewhere (Kelly 2002: 172-4). If anything, I think the situation Kelly brings up is even less puzzling than these already unpuzzling cases. After all, part of the point of modestly moderating one's views in response to peer disagreement is to mitigate the damaging effects of one's fallibility. So it shouldn't be surprising if correctly doing so sometimes brings imperfect agents to credences that less imperfect agents would have also arrived at. If anything, that is the equal weight principle doing exactly its job!

[377] Notice also, as a precondition of this, that the requirement does refer to one's *believing* one's disputant to be a peer, and not to the disputant *actually* being one's peer – a difference that is not always carefully attended to in the disagreement literature. Once we are construing the requirement as a coherence requirement, this is the obvious way to go. Again, because the requirement is wide-scope, this does not have the bad result that when one makes unreasonable evaluations of one's disputant's reliability, these unreasonable evaluations have some kind of fixed authority over what one ought to believe.

[378] See Christensen (2011); Elga (2007: esp. 487); White (2009: 247-9).

Mixed-scope equal weight principle. If one believes that some other

doxastic agent A is one's peer, then rationality requires that if one has credence

C1 for proposition p, and one learns that A has credence C2 for proposition

p, then one revises one's credence to ½(C1+C2).

I've called this version of the principle "mixed-scope" because even though the belief in

peerhood is now outside the scope of the 'requires' operator, the 'requires' operator still takes

wide-scope over the conditional that remains. This principle would still allow us to resist

Kelly's objection. However, I myself disagree with the original conciliationists, and prefer the

pure wide-scope principle to the mixed-scope principle. This is because I think that it is often

a reasonable response to disagreement to downgrade one's estimate of one's disputant's

reliability.[379] In other work (Worsnip 2014), I have defended this claim, and tried to say what

makes the difference as to when one should revise one's credence and when one should revise

one's estimate of one's disputant's reliability (or, more accurately, by how much, since often I

think it is most reasonable to revise both to some extent). But the point of the relevant

coherence requirement, I think, is that it says that the two must be brought into a kind of

equilibrium: it's irrational to revise *neither*.[380] This is because it's incoherent to think that your

disputant is as reliable as you about the matter at hand, but that you are far more likely to be

---

[379] Do I fail to count as a conciliationist, then? My view is a compromise view. Unlike conciliationist views, it allows for frequently revising one's estimate of one's disputant's reliability. Unlike "right reasons" or "total evidence" views, it says that in and of itself disagreement rationally requires revision in one's doxastic states – and that what is most reasonable often, even usually, involves revising one's credence. For much more see Worsnip (2014). There I argue that when the two traditional views are adequately moderated to take account of their obvious flaws, and when we make some relevant distinctions that create a false appearance of division, they converge on the compromise view that I argue for.

[380] Wiland (ms.) has independently developed a view very similar to mine in this regard, and explicitly notes its wide-scope character.

right (or to think that you are equally likely to be right, but still to hold unwaveringly to your own view).

Ultimately, I do think that the wide-scope equal weight principle, as stated above, is too simple to capture the coherence requirement in its entirety. In particular, an on/off notion of someone's being one's peer or not is too coarse. There is a spectrum of reliability-evaluations one can make with respect to one's disputant. Moreover, as I said, it is often reasonable to downgrade both one's estimate of one's disputant's reliability and one's credence. Coherence puts constraints on how the two must be adjusted relative to one another. The limiting case is that in which one's estimate of one's disputant's reliability does not change at all, and in this case one's new credence should be equidistant between the two original credences, as the requirement says. But this is just the limiting case. If one only slightly downgrades one's estimate of one's peer's reliability, one still needs to make a significant adjustment to one's credence (albeit not quite one that gives one's disputant's view equal weight) to be coherent. Again, there are many questions for future work here, and I have begun to answer them in Worsnip (2014). But these details do not matter here, for the fundamental illustration of how construing the relevant requirement as a wide-scope coherence requirement helps in the debate about peer disagreement.

We also face the same question that we faced in the discussion of conditionalization above about whether the requirement is diachronic or synchronic. As with one's priors, resolving this requires saying whether we interpret the idea of one's credence being "original" or "pre-disagreement" in temporal fashion or in some non-temporal fashion. Even if we do think of this pre-disagreement credence in temporal fashion, one's post-disagreement credence and estimate of one's interlocutor's reliability (or belief in her peerhood) should be thought of as synchronous. So in this case, I am more confident that the requirement as a

whole retains significant features like those of the synchronic requirements we encountered earlier in this dissertation. Again, however, there is clearly more to be said here.

(c) Other directions for future work

In addition to investigating accounts of conditionalization and disagreement principles as wide-scope coherence requirements, there are a number of issues in and around this dissertation that merit revisiting. I will mention just four. First, I have been noticeably silent on whether it is a requirement of rationality in the sense I'm interested in that one's credences be probabilistically coherent.[381] This is a very demanding requirement, in contrast to many of the requirements that I have been considering. On the other hand, it is also clearly a candidate coherence constraint in the sense that it is governs how one combines various different mental states. I am genuinely unsure as to whether it should be accepted as part of the account I defend. If it is, there are no doubt interesting questions about how it interacts with many of the other claims defended here.

Second, there is a lot more to be said about the ideas that I offered very tentatively in section 2.1 and 2.2 about the metaethics of coherence requirements. In particular, I think the idea that the metaethics for coherence requirements differs considerably from that for substantive reasons – perhaps in its being more amenable to a naturalistic treatment – is intriguing and worth exploring further. I hope to pursue this elsewhere.

---

[381] I said in chapter 5 that it is permissible not to radically deviate from probabilistic coherence. But that is a much weaker claim.

Third, I hope to revisit whether there is some way of getting the coherence requirement that captures the impermissibility of *small* (but larger than 2-belief) sets of inconsistent beliefs (see section 5.9 above).

Last, and most ambitiously, I hope to pursue the role that rationality-attributions play in social science. As I mentioned in section 2.1, because coherence requirements play an explanatory role that substantive reasons do not, a notion of rationality that gives pride of place to coherence is well-suited for use in social science. Yet many social scientists have misdescribed their own assumed theories of rationality as giving a starring role to self-interest. Philosophers have tended take these claims at face value, and have rightly been suspicious of the self-interest theory of rationality that they take to be at the heart of social science (especially economics). But I suspect that once we adequately clarify the relationship between a tightly related bundle of notions including utility, preference and well-being, we can see that the notion of rationality truly at the heart of economic theory is one of rationality as coherence, not one of rationality as self-interest in the ordinary sense. Verifying this is a major hermeneutic project to which I would like to make a contribution. My hope is that, by embracing the notion of rationality as coherence, we can arrive at a notion that has a chance at unifying the philosophical and social-scientific study of rationality.

(d) Rationality as a human trait

Some philosophers find the distinction between coherence requirements and reasons a bit esoteric, or a boring exercise in taxonomy. Part of what I have been trying to do in this dissertation is to show that the distinction really matters for epistemology and for philosophy

315

more broadly. I have been trying to show that coherence requirements are a distinctive, interesting, and important part of the landscape of doxastic normativity, broadly construed.

As I suggested in section 2.1, coherence is central to our understanding of human agency. We are disposed to generally comply with coherence requirements – since violations of such requirements are hard to sustain under conditions where one's mental states are fully transparent to oneself. Indeed, it is plausibly part of what it is to be in a particular mental state that one is disposed to comply with the coherence requirements that govern it, at least under such transparent conditions.

As long as one is thinking of rationality in terms of responding to substantive normative reasons, it looks like rationality isn't something that we are even approximately disposed to approach, nor something that can play much of a role in explaining our behavior. Beliefs against the evidence are widespread. We can count the ways in which we are non-ideal *ad nauseum*. But to simply conclude from this that humans are irreparably irrational is to elide a core sense in which human beings *are* rational: they are disposed, at least generally, to satisfy the coherence requirements. That may not be the virtue of always doing and believing what we ought. But in its own way, it is a remarkable thing. If we let coherence requirements be obscured entirely by substantive reasons, we miss out on the contemplation of one of the most magnificent, distinctive, and explanatorily important features of what it is to be human.

References

Adler, J.E. (2002a). 'Akratic Believing?,' *Philosophical Studies*, 110/1: 1-27.

-------- (2002b). *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Alston, W.P. (1980). 'Level Confusions in Epistemology,' *Midwest Studies in Philosophy*, 5/1: 135-150.

Anderson, E. (2001). 'Unstrapping the Straitjacket of 'Preference',' *Economics & Philosophy*, 17: 21-38.

Audi, R. (2001). *The Architecture of Reason: The Structure and Substance of Rationality*. Oxford: Oxford University Press.

Ballantyne, N. & Coffman, E.J. (2011). 'Uniqueness, Evidence, and Rationality,' *Philosophers' Imprint*, 11/18.

Bergmann, M. (2005). 'Defeaters and Higher-Level Requirements,' *Philosophical Quarterly*, 55/220: 419-436.

Björnsson, G. & Finlay, S. (2010). 'Metaethical Contextualism Defended,' *Ethics*, 121/1: 7-36.

Blackburn, S. (1995). 'Practical Tortoise Raising,' *Mind*, 104/416: 695-711.

-------- (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Bovens, L. (1995). 'The Intentional Acquisition of Mental States', *Philosophy and Phenomenological Research*, 55/4: 821-840.

Bratman, M.E. (1987). *Intention, Plans and Practical Reason*. Cambridge, MA: Harvard University Press.

-------- (2009). 'Intention, Practical Rationality, and Self-Governance,' *Ethics*, 119/3: 411-443.

Broome, J. (1991). ''Utility','' *Economics & Philosophy*, 7/1: 1-12.

-------- (1999). 'Normative requirements,' *Ratio*, 12: 398-419.

-------- (2004). 'Reasons,' in Wallace, Pettit & Scheffler (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press.

-------- (2007). 'Wide or Narrow Scope?,' *Mind*, 116/462: 359-370.

-------- (2013). *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.

Brown, J. (2015). 'Evidence and Epistemic Evaluation,' *Oxford Studies in Epistemology,* 5.

Brunero, J. (2010). 'The scope of rational requirements,' *Philosophical Quarterly*, 60/238: 28-49.

-------- (2012). 'Instrumental rationality, symmetry, and scope,' *Philosophical Studies*, 157/1: 125-140.

Buchak, L. (2012). 'Can it be rational to have faith?,' in Chandler & Harrison (eds.), *Probability in the Philosophy of Religion*. Oxford: Oxford University Press.

-------- (2014). 'Belief, credence, and norms,' *Philosophical Studies*, 169: 285-311.

Cariani, F., Kaufmann, M. & Kaufmann, S. (2013). 'Deliberative Modality under Epistemic Uncertainty,' *Linguistics and Philosophy* 36/3: 225–59.

Carroll, L. (1895). 'What The Tortoise Said To Achilles,' *Mind*, 4: 278-80.

Christensen, D. (2004). *Putting Logic in its Place*. Oxford: Oxford University Press.

-------- (2007). 'Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals,' *Oxford Studies in Epistemology*, 2: 3-31.

-------- (2010a). 'Higher-Order Evidence,' *Philosophy and Phenomenological Research*, 81/1: 185-215.

-------- (2010b). 'Rational Reflection,' *Philosophical Perspectives*, 24: 121-140.

-------- (2011). "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosophers' Imprint*, 11/6.

-------- (2013). 'Epistemic Modesty Defended,' in Christensen & Lackey (2013).

Christensen, D. & Lackey, J. (eds.) (2013). *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Clarke, R. (2013). 'Belief is Credence One (in Context),' *Philosophers' Imprint*, 13/11: 1-18.

-------- (forthcoming). 'Preface Writers are Consistent,' *Pacific Philosophical Quarterly*. Page references are to the draft at http://www.rogerclarke.org/research/Preface%20for%20PPQ%20-%20submitted.pdf?attredirects=0.

Coates, A. (2012). 'Rational Epistemic Akrasia,' *American Philosophical Quarterly*, 49/2: 113-124.

Cohen, S. (forthcoming). 'Theorizing About the Epistemic,' *Inquiry*.

Conee, E. & Feldman, R. (1985). 'Evidentialism,' *Philosophical Studies*, 48/1: 15-34.

Copp, D. (2007). *Morality in a Natural World*. Cambridge, UK: Cambridge University Press.

Dancy, J. (1977). 'The logical conscience,' *Analysis*, 37/2: 81-84.

-------- (2000). *Practical Reality*. Oxford: Oxford University Press.

Darwall, S. (1983). *Impartial Reason*. Ithaca: Cornell University Press.

Davidson, D. (2004). *Problems of Rationality*. Oxford: Oxford University Press.

Dennett, D. (1971). 'Intentional Systems,' *Journal of Philosophy*, 68/4: 87-106.

DeRose, K. (2000). 'Ought We to Follow Our Evidence?,' *Philosophy and Phenomenological Research*, 60/3: 697-706.

-------- (2009). *The Case for Contextualism*. Oxford: Oxford University Press.

-------- (forthcoming). *Knowledge, Skepticism & Context, Vol. 2*. Oxford: Oxford University Press.

Dowell, J. (2012). 'Contextualist solutions to three puzzles about practical conditionals,' *Oxford Studies in Metaethics*, 7: 271-303.

Dretske, F. (2005). 'The Case Against Closure,' in Steup & Sosa (eds.), *Contemporary Debates in Epistemology*. Oxford: Blackwell.

Easwaran, K. & Fitelson, B. (2015). 'Accuracy, Coherence, and Evidence,' *Oxford Studies in Epistemology*, 5.

Elga, A. (2005). 'On Overrating Oneself…and Knowing It,' *Philosophical Studies*, 133/1-2: 115-124.

Elga, A. & Rayo, A. (ms.). 'Fragmentation and information access.' Draft manuscript, MIT/Princeton.

Enoch, D. (2006). 'Agency, Shmagency: Why Normativity Won't Come From What Is Constitutive of Action,' *Philosophical Review*, 115/2: 169-198.

Evnine, S.J. (1999). 'Believing Conjunctions,' *Synthese*, 118/2: 201-227.

Fantl, J. & McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford: Oxford University Press.

Feldman, R. (1988). 'Subjective and Objective Justification in Ethics and Epistemology,' *The Monist*, 71/3: 405-419.

-------- (2000). 'The Ethics of Belief,' *Philosophy and Phenomenological Research*, 60/3: 667-695.

-------- (2005). 'Respecting the Evidence,' *Philosophical Perspectives*, 19: 95-119.

Fitelson, B. (2012). 'Evidence of evidence is not (necessarily) evidence,' *Analysis*, 72/1: 85-88.

Finlay, S. (2010). 'What *ought* probably means, and why you can't detach it,' *Synthese*, 177: 67-89

Fogal, D. (ms.). 'Rationality and Requirements,' draft manuscript, New York University.

Foley, R. (1993). *Working without a Net*. Oxford: Oxford University Press.

Foot, P. (1972). 'Morality as a System of Hypothetical Imperatives,' *Philosophical Review*, 81/3: 305-316.

Forrester, J.W. (1984). 'Gentle Murder, or the Adverbial Samaritan,' *Journal of Philosophy*, 81: 193-196.

van Fraassen, B.C. (1972). 'The Logic of Conditional Obligation,' *Journal of Philosophical Logic*, 1: 417-438.

-------- (1995). 'Fine-Grained Opinion, Probability, and the Logic of Full Belief,' *Journal of Philosophical Logic*, 24/4: 349-377.

Ganson, D. (2008). 'Evidentialism and pragmatic constraints on outright belief,' *Philosophical Studies*, 139: 441-458.

Gendler, T.S. (2002). 'Personal Identity and Thought-Experiments,' *The Philosophical Quarterly*, 52/206: 34-54.

-------- (2008a). 'Alief and Belief,' *Journal of Philosophy*, 105/10: 634-663.

-------- (2008b). 'Alief in Action (and Reaction),' *Mind & Language*, 23/5: 552-85.

-------- (2011). 'On The Epistemic Costs of Implicit Bias,' *Philosophical Studies*, 156: 33-63.

Gensler, H.J. (1985). 'Ethical Consistency Principles,' *Philosophical Quarterly*, 35/139: 156-170.

Gibbard, A. (2005). 'Truth and Correct Belief,' *Philosophical Issues*, 15: 338-350.

Gibbons, J. (2010). 'Things that makes things reasonable,' *Philosophy and Phenomenological Research*, 81/2: 335-61.

Gillies, A. (2001). 'A New Solution to Moore's Paradox,' *Philosophical Studies*, 105: 237-50.

Goldman, A. (1976). 'Discrimination and Perceptual Knowledge,' *Journal of Philosophy*, 73: 771-791.

-------- (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

-------- (2009). 'Internalism, Externalism, and the Architecture of Justification,' *Journal of Philosophy*, 106/6: 309-338.

Greco, D. (2013). 'Probability and Prodigality,' *Oxford Studies in Epistemology*, 3: 82-107.

-------- (2014a). 'A puzzle about epistemic akrasia,' *Philosophical Studies*, 167/2: 201-219.

-------- (2014b). 'Could KK be OK?,' *Journal of Philosophy*, 111/4: 169-197.

-------- (2014c). 'Iteration and Fragmentation,' *Philosophy and Phenomenological Research*, Online Early View.

Greenspan, P.S. (1975). 'Conditional Oughts and Hypothetical Imperatives,' *Journal of Philosophy*, 72/10: 259-276.

Hampshire, S. (1965). *Freedom of the Individual*. New York: Harper & Row.

Hansson, B. (1969). 'An Analysis of Some Deontic Logics,' *Noûs*, 3: 373-398.

Hare, R.M. (1971). 'Wanting: Some Pitfalls,' in Binkley, Bronaugh & Marias (eds.), *Agent, Action, and Reason*. Toronto: University of Toronto Press.

Harman, G. (1986). *Change in View*. Cambridge, MA: MIT Press.

-------- (1999). *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.

Hawthorne, James & Bovens, L. (1999). 'The Preface, the Lottery, and the Logic of Belief,' *Mind*, 109/430: 241-264.

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

Hawthorne, John & Srinivasan, A. (2013). 'Disagreement Without Transparency: Some Bleak Thoughts,' in Christensen & Lackey (2013).

Hazlett, A. (2012). 'Higher-Order Epistemic Attitudes and Intellectual Humility,' *Episteme*, 9/3: 205-223.

Hedden, B. (forthcoming). 'Time-Slice Rationality,' *Mind*.

Henning, T. (2014). 'Normative Reasons Contextualism,' *Philosophy and Phenomenological Research*, 88/3: 593-624.

Hieronymi, P. (2005). 'The Wrong Kind of Reason,' *Journal of Philosophy*, 102/9: 437-57.

------- (2006). 'Controlling Attitudes,' *Pacific Philosophical Quarterly*, 87/1: 45-74.

------- (2009). 'Believing at Will,' *Canadian Journal of Philosophy*, Supplementary Volume 35: 149-187.

Hill, T.E. (1973). 'The Hypothetical Imperative,' *Philosophical Review*, 82/4: 429-450.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca: Cornell University Press.

Horowitz, S. (2014). 'Epistemic Akrasia,' *Noûs*, 48/4: 718-744.

Huemer, M. (2011). 'The Puzzle of Metacoherence,' *Philosophy and Phenomenological Research*, 82/1: 1-21.

Hurley, S.L. (1989). *Natural Reasons*. Oxford: Oxford University Press.

Hussain, N.J.Z. (ms.). 'The Requirements of Rationality,' draft manuscript, Stanford University. Page references are to the draft available at http://web.stanford.edu/~hussainn/StanfordPersonal/Online_Papers_files/HussainRequirementsv24.pdf.

Jackson, F. (1991). 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection,' *Ethics*, 101/3, 461-482.

James, W. (1896). 'The Will to Believe,' *The New World*.

Jeffrey, R. (2004). *Subjective Probability: The Real Thing*. Cambridge, UK: Cambridge University Press.

Joyce, J. (1998). 'A nonpragmatic vindication of probabilism,' *Philosophy of Science*, 65: 575-603.

-------- (2005). 'How probabilities reflect evidence,' *Philosophical Perspectives*, 19: 154-178.

Kamm, F.M. (2007). *Intricate Ethics*. Oxford: Oxford University Press.

Kaplan, M. (1996). *Decision Theory as Philosophy*. Cambridge, UK: Cambridge University Press.

Keller, S. (2004). 'Friendship and Belief,' *Philosophical Papers*, 33/3: 329-351.

Kelly, T. (2002). 'The Rationality of Belief and Some Other Propositional Attitudes,' *Philosophical Studies*, 110/2: 163-96.

-------- (2003). 'Epistemic Rationality as Instrumental Rationality: A Critique,' *Philosophy and Phenomenological Research*, 66/3: 612-640.

-------- (2005). "The Epistemic Significance of Disagreement," *Oxford Studies in Epistemology*, 1: 167-196.

-------- (2006). 'Evidence,' *Stanford Encyclopedia of Philosophy*.

-------- (2007). 'Evidence and Normativity: Reply to Leite,' *Philosophy and Phenomenological Research*, 75/2: 465-474.

-------- (2010). 'Peer Disagreement and Higher Order Evidence,' in Feldman & Warfield (eds.), *Disagreement*. Oxford: Oxford University Press.

Kiesewetter, B. (ms.-a). *The Normativity of Rationality*. Dissertation, Humboldt University Berlin.

-------- (ms.-b). 'You Ought to Φ Only If You May Believe That You Ought to Φ,' draft manuscript, Humboldt University Berlin.

Kolodny, N. (2005). 'Why be rational?,' *Mind*, 114/455: 509-563.

-------- (2007). 'How does coherence matter?,' *Proceedings of the Aristotelian Society*, 107/3: 229-263.

-------- (2008a). 'Why Be Disposed to Be Coherent?,' *Ethics*, 118/3: 437-463.

-------- (2008b). 'The Myth of Practical Consistency,' *European Journal of Philosophy*, 16/3: 366-402.

Kolodny, N. & MacFarlane, J. (2010). 'Ifs and Oughts,' *Journal of Philosophy*, 107/3: 115-143.

Kornblith, H. (1993). 'Epistemic Normativity,' *Synthese*, 94/3: 357-376.

Korsgaard, C.M. (2008). 'The Normativity of Instrumental Reason,' in her *The Constitution of Agency*. Oxford: Oxford University Press.

-------- (2009). *Self-Constitution: Agency, Identity & Integrity*. Oxford: Oxford University Press.

Kratzer, A. (1981). 'The Notional Category of Modality,' in Eikmeyer & Rieser (eds.), *Words, Worlds, and Contexts*. Berlin: de Gruyter.

-------- (2012). *Modals and Conditionals*. Oxford: Oxford University Press.

Lackey, J. (2010). 'A Justificationist View of Disagreement's Epistemic Significance,' in Haddock, Millar & Pritchard (eds.), *Social Epistemology*. Oxford: Oxford University Press.

Lasonen-Aarnio, M. (2008). 'Single premise deduction and risk,' *Philosophical Studies*, 141/2: 157-173.

-------- (2014). 'Higher-order evidence and the limits of defeat,' *Philosophy and Phenomenological Research*.

-------- (2015). 'New Rational Reflection and Internalism about Rationality,' *Oxford Studies in Epistemology*, 5.

Leitgeb, H. (2014a). 'A way out of the preface paradox?,' *Analysis*, 74/1: 11-15.

-------- (2014b). 'The stability theory of belief,' *Philosophical Review*, 123/2: 131-171.

Levi, I. (1983). *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

-------- (1996). 'Elusive Knowledge,' *Australasian Journal of Philosophy*, 74/4: 549-567.

Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge, UK: Cambridge University Press.

Lord, E. (2010). 'Having reasons and the factoring account,' *Philosophical Studies*, 149: 283-296.

-------- (2011). 'Violating Requirements, Exiting From Requirements, and the Scope of Rationality,' *Philosophical Quarterly*, 61/243: 392-399.

-------- (2014a). 'The real symmetry problem(s) for wide-scope accounts of rationality,' *Philosophical Studies*, 170/3, 443-464.

-------- (2014b). 'The coherent and the rational,' *Analytic Philosophy*, 55/2, 151-175.

-------- (ms.). 'What You're Rationally Required to Do and What You Ought to Do (Are The Same Thing!),' draft manuscript, University of Pennsylvania.

Mackinson, D.C. (1965). 'The Paradox of the Preface,' *Analysis*, 25/6: 205-207.

McDowell, J. (1998). *Mind, Value & Reality*. Cambridge, MA: Harvard University Press.

Meiland, J.W. (1980). 'What Ought We to Believe? Or The Ethics of Belief Revisited,' *American Philosophical Quarterly*, 17/1: 15-24.

Murray, S.E. (2014). 'Varieties of Update,' *Semantics & Pragmatics*, 7/2: 1-53.

Nagel, J. (2008). 'Knowledge ascriptions and the psychological consequences of changing stakes,' *Australasian Journal of Philosophy*, 86/2: 279-294.

Neta, R. (2003). 'Contextualism and the Problem of the External World,' *Philosophy and Phenomenological Research*, 66/1: 1-31.

Norby, A. (2014). 'Against Fragmentation,' *Thought*, 3/1: 30-38.

Nozick, R. (1993). *The Nature of Rationality*. Princeton: Princeton University Press.

Olson, J. (2004). 'Buck-Passing and the Wrong Kind of Reasons,' *Philosophical Quarterly*, 54/215: 295-300.

Owens, D. (2000). *Reason Without Freedom: The Problem of Epistemic Normativity*. London: Routledge.

Parfit, D. (2011). *On What Matters: Vol. 1*. Oxford: Oxford University Press.

Pittard, J. & Worsnip, A. (ms.). 'Metanormative Contextualism and Normative Uncertainty,' draft manuscript, Yale University.

Pettit, P. & Smith, M. (1996). 'Freedom in Belief and Desire,' *Journal of Philosophy*, 93: 433-36.

Piller, C. (2006). 'Content-Related and Attitude-Related Reasons for Preferences,' *Royal Institute of Philosophy Supplement*, 59: 155-182.

Plantinga, A. (1981). 'Is Belief in God Properly Basic?,' *Noûs*, 15: 41-51.

Price, H.H. (1973). *Perception*. London: Methuen.

Pritchard, D. (2002). 'Radical Scepticism, Epistemological Externalism, and Closure,' *Theoria*, 68/2: 129-161.

Pryor, J. (2004). 'What's Wrong with Moore's Argument?,' *Philosophical Issues*, 14: 349-378.

-------- (2013). 'Problems for Credulism,' in Tucker (ed.), *Seemings and Justification*. Oxford: Oxford University Press.

Radford, C. (1966). 'Knowledge – By Examples,' *Analysis*, 27: 1-11.

Railton, P. (1986). 'Moral Realism,' *Philosophical Review*, 95/2: 163-207.

-------- (1994). 'Truth, Reason, and the Regulation of Belief,' *Philosophical Issues*, 5: 71-93.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Raz, J. (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.

Reisner, A. (2009). 'The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem,' *Philosophical Studies*, 145: 257-272.

Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.

Rippon, S. (2011). 'In Defense of the Wide-Scope Instrumental Principle,' *Journal of Ethics and Social Philosophy*, 5/2: 1-21.

Ross, J. & Schroeder, M. (2014). 'Belief, Credence, and Pragmatic Encroachment,' *Philosophy and Phenomenological Research*, 88/2: 259-288.

Roush, S. (2010). 'Optimism about the Pessimistic Induction,' in Magnus & Busch (eds.), *New Waves in Philosophy of Science*. New York: Palgrave-MacMillan.

Scanlon, T.M. (1998). *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.

-------- (2004). 'Reasons: A Puzzling Duality?,' in Wallace, Pettit & Scheffler (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press.

-------- (2007). 'Structural Irrationality,' in Brennan, Goodin, Jackson & Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford: Oxford University Press.

Schechter, J. (2013). 'Rational Self-Doubt and the Failure of Closure,' *Philosophical Studies*, 163/2: 428-452.

Schoenfield, M. (2014). 'Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief,' *Noûs*, 48/2: 193-218.

Schroeder, M. (2004). 'The Scope of Instrumental Reason,' *Philosophical Perspectives,* 18: 337-364.

-------- (2008). 'Having reasons,' *Philosophical Studies*, 139: 57-71.

-------- (2009). 'Means-end coherence, stringency, and subjective reasons,' *Philosophical Studies*, 143: 223-248.

-------- (2012). 'The Ubiquity of State-Given Reasons,' *Ethics*, 122/3: 457-488.

Schwitzgebel, E. (2002). 'A Phenomenal, Dispositional Account of Belief'. *Noûs*, 36/2: 249-75.

-------- (2008). 'The Unreliability of Naïve Introspection,' *Philosophical Review*, 117/2: 245-73.

-------- (2010). 'Acting Contrary to our Professed Beliefs, or, The Gulf Between Occurrent Judgment and Dispositional Belief,' *Pacific Philosophical Quarterly*, 91: 531-553.

Sen, A. (1973). 'Behaviour and the Concept of Preference,' *Economica*, 40/159: 241-259.

Sepielli, A. (2014). 'What to Do When You Don't Know What to Do When You Don't Know What to Do…,' *Noûs*, 48/3: 521-544.

Setiya, K. (2004). 'Against Internalism,' *Noûs,* 38/2: 266-298.

-------- (2007). 'Cognitivism About Instrumental Reason,' *Ethics*, 117/4: 649-673.

-------- (2008). 'Believing at Will,' *Midwest Studies in Philosophy*, 32: 36-52.

Shah, N. (2003). 'How Truth Governs Belief,' *Philosophical Review*, 112/4: 447-482.

-------- (2006). 'A New Argument for Evidentialism,' *Philosophical Quarterly*, 56/225: 481-498.

Shah, N. & Velleman, J.D. (2005). 'Doxastic Deliberation,' *Philosophical Review*, 114/4: 497-534.

Sherman, B. & Harman, G. (2011). 'Knowledge and Assumptions,' *Philosophical Studies*, 156/1: 131-140.

Shpall, S. (2013). 'Wide and Narrow Scope,' *Philosophical Studies*, 163/3: 717-736.

Silk, A. (2014). 'Why 'Ought' Detaches: Or, Why You Ought to Get With My Friends (If You Want to Be My Lover),' *Philosophers' Imprint*, 14/7.

Skorupski, J. (2010). *The Domain of Reasons*. Oxford: Oxford University Press.

Smith, Martin. (2010). 'What Else Justification Could Be,' *Noûs*, 44/1: 10-31.

Smith, Michael. (2007). 'Is There A Nexus Between Reasons and Rationality?,' in Tenenbaum (ed.), *Moral Psychology*. Amsterdam: Rodopi.

Smithies, D. (2012). 'Moore's Paradox and the Accessibility of Justification,' *Philosophy and Phenomenological Research*, 85/2: 273-300.

Southwood, N. (2008). 'Vindicating the normativity of rationality,' *Ethics*, 119: 9-30.

-------- (ms.). 'The normativity of rationality without reasons,' draft manuscript, Australian National University.

Srinivasan, A. (2015). 'Are We Luminous?,' *Philosophy and Phenomenological Research*, 90/2: 294-319.

-------- (forthcoming). 'Normativity without Cartesian Privilege,' *Philosophical Issues*.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

-------- (1999). *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

-------- (2002). 'Common Ground,' *Linguistics & Philosophy*, 25: 701-21.

Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Sterba, J.P. (2013). *From Rationality to Equality*. Oxford: Oxford University Press.

Stroud, S. (2006). 'Epistemic Partiality in Friendship,' *Ethics*, 116/3: 498-524.

Sturgeon, S. (2008). 'Reason and the Grain of Belief,' *Noûs*, 42/1: 139-165.

Sutton, J. (2007). *Without Justification*. Cambridge, MA: MIT Press.

Sylvan, K. (ms.). 'On Divorcing the Rational and the Justified in Epistemology,' draft manuscript, University of Southampton. Page references are to the draft manuscript available at https://www.academia.edu/4773704/On_Divorcing_the_Rational_and_the_Justified_in_E pistemology.

Titelbaum, M.G. (2010). 'Not Enough There There: Evidence, Reasons, and Language Independence,' *Philosophical Perspectives*, 24: 477-528.

-------- (2015a). 'How to derive a narrow-scope requirement from wide-scope requirements,' *Philosophical Studies*, 172/2: 535-542.

-------- (2015b). 'Rationality's Fixed Point (or: In Defense of Right Reason),' *Oxford Studies in Epistemology*, 5.

Unger, P. (1986). 'The Cone Model of Knowledge,' *Philosophical Topics*, 14: 125-178.

Velleman, J.D. (1989). *Practical Reflection*. Princeton: Princeton University Press.

-------- (2000). 'On The Aim of Belief', in *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Veltman, F. (1996). 'Defaults in Update Semantics,' *Journal of Philosophical Logic*, 25: 221-61.

Wallace, R.J. (2001). 'Normativity, Commitment, and Instrumental Reason,' *Philosophers' Imprint*, 1/4.

Way, J. (2010). 'Defending the wide-scope approach to instrumental reason,' *Philosophical Studies*, 147/2: 213-233.

-------- (2011). 'The symmetry of rational requirements,' *Philosophical Studies*, 155/2: 227-239.

Weatherson, B. (2005). 'Can We Do Without Pragmatic Encroachment?,' *Philosophical Perspectives*, 19: 417-443.

-------- (2012). 'Knowledge, Bets, and Interests', in Brown & Gerken (eds), *Knowledge Ascriptions*. Oxford: Oxford University Press.

-------- (ms.) 'Do Judgments Screen Evidence?,' draft manuscript, University of Michigan.

Wedgwood, R. (2002a). 'Internalism Explained,' *Philosophy and Phenomenological Research*, 65/2: 349-69.

-------- (2002b). 'The Aim of Belief,' *Philosophical Perspectives*, 16: 267-297.

-------- (2003). 'Choosing Rationally and Choosing Correctly,' in Stroud & Tappolet (eds.), *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press.

-------- (2006). 'The Meaning of 'Ought',' *Oxford Studies in Metaethics*, 1: 127-160.

-------- (2007). 'A Simple Solution to the "Preface Paradox"'. Blog post at *Certain Doubts*, dated April 20th, 2007.

-------- (2012a). 'Justified inference,' *Synthese*, 189/2: 273-295.

-------- (2012b). 'Outright Belief,' *dialectica*, 66/3: 309-329.

White, R. (2005). 'Epistemic Permissiveness,' *Philosophical Perspectives*, 19: 445-459.

-------- (2007). 'Epistemic Subjectivism,' *Episteme*, 4/1: 115-129.

White, R. (2009). 'On Treating Oneself and Others as Thermometers,' *Episteme*, 6/3: 233-250.

Whiting, D. & Way, J (ms.). 'If You Justifiably Believe That You Ought to Φ, You Ought to Φ,' draft manuscript, University of Southampton.

Wiland, E. (ms.). 'The Wide-Scoped Total Evidence View,' draft manuscript, University of Missouri-St Louis.

Williams, B.A.O. (1973). 'Deciding to Believe,' in his *Problems of the Self*. Cambridge, UK: Cambridge University Press.

-------- (1981). 'Internal and external reasons,' in his *Moral Luck*. Cambridge, UK: Cambridge University Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

-------- (2005). 'Knowledge and Scepticism,' in Jackson & Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press.

-------- (2009). 'Reply to John Hawthorne and Maria Lasonen-Aarnio,' in Greenough & Pritchard (eds.), *Williamson on Knowledge*. Oxford: Oxford University Press.

-------- (2011). 'Improbable Knowing,' in Dougherty (ed.), *Evidentialism and its Discontents*. Oxford: Oxford University Press.

Woods, M. (1972). 'Reasons for Action and Desires,' *Proceedings of the Aristotelian Society Supplementary Volume*, 46: 189-201.

Worsnip, A. (2014). 'Disagreement about Disagreement? What Disagreement about Disagreement?,' *Philosophers' Imprint*, 14/18.

-------- (forthcoming-a). 'Possibly False Knowledge,' *Journal of Philosophy*.

-------- (forthcoming-b). 'Two Kinds of Stakes,' *Pacific Philosophical Quarterly*.

-------- (ms.). 'Belief, Possible Worlds, and the Preface Paradox,' draft manuscript, Yale University.

Wright, C. (2004). 'Warrant for Nothing (and Foundations for Free)?,' *Proceedings of the Aristotelian Society Supplementary Volume*, 78: 167-212.