IN DEFENSE OF RIGHT REASON

MICHAEL G. TITELBAUM

Draft of August 16, 2012.
Please do not circulate or cite without permission.

Rational requirements have a special status in the theory of rationality. This is obvious in one sense: they supply the *content* of that theory. But I want to suggest that rational requirements have another special status—as *objects* of the theory of rationality. In slogan form, my thesis is:

> **Fixed Point Thesis:** Mistakes *about* the requirements of rationality are mistakes *of* rationality.

The key claim in the Fixed Point Thesis is that the mistakes in question are *rational* mistakes. If I incorrectly believe that something is a rational requirement, I clearly have made a mistake in some sense, in that I have a false belief. But in many cases possession of a false belief does not indicate a *rational* mistake; when evidence is misleading, one can rationally believe a falsehood. According to the Fixed Point Thesis, this cannot happen with beliefs about the requirements of rationality—any false belief about the requirements of rationality involves a mistake not only in the sense of believing something *false* but also in a distinctly *rational* sense. While the Fixed Point Thesis is a claim about theoretical rationality (it concerns what we are rationally permitted to *believe*), it applies both to mistakes about the requirements of theoretical rationality and to mistakes about requirements of practical rationality.

Like any good philosophical slogan, the Fixed Point Thesis requires qualification. Suppose I falsely believe that what Frank just wrote on a napkin is a requirement of rationality, because I am misled about what exactly Frank wrote. In some sense my false belief is about the requirements of rationality, but I need not have made a rational mistake. This suggests that the Fixed Point Thesis should be restricted to mistakes involving *a priori* rational-requirement truths. So from now on when I discuss beliefs about rational requirements I will be considering only beliefs in *a priori* truths or falsehoods.[1] It may be that the set of beliefs about rational requirements targeted by the Fixed Point Thesis should be restricted even more than that. As I build my case for the thesis, we'll see how far we can make it extend.

Even restricted to *a priori* rational-requirement beliefs (or a subset thereof), the Fixed Point Thesis is surprising—if not downright incredible. As I understand it, rationality concerns constraints on practical and theoretical reasoning arising from consistency requirements among an agent's attitudes, evidence, and whatever else reasoning takes into account.[2] One does not expect such consistency requirements

---

[1]By an "*a priori* truth" I mean something that can be known *a priori*, and by an "*a priori* falsehood" I mean the negation of an *a priori* truth.

[2]While some may want to use the word "rationality" in a more externalist way, I take it most of us recognize at least some normative notion meeting the description just provided (whatever

to specify specific contents it is irrational to believe. While there have long been those (most famously, Kant) who argue that practical rationality places specific, substantive requirements on our intentions and/or actions, one rarely sees arguments for such substantive rational requirements on belief.[3] Moreover, the Fixed Point Thesis has the surprising consequence (as I'll explain later) that one can never have all-things-considered misleading evidence about rational requirements.

Finally, the Fixed Point Thesis has consequences for a topic of much recent controversy: peer disagreement. Given a case in which an agent discovers she has drawn the opposite conclusion from an equally-rational peer in response to the same body of evidence, the most initially appealing position (called Split the Difference) holds that the agent should suspend judgment about the matter in question in response to discovering the disagreement. Yet the Fixed Point Thesis supports the Right Reasons position, holding that whichever peer originally drew the conclusion actually supported by her evidence should not abandon that conclusion after discovering the disagreement.

Despite both its initial implausibility and its unexpected consequences, we can argue to the Fixed Point Thesis from a premise most of us accept already: that *akrasia* is irrational. After connecting the Fixed Point Thesis to logical omniscience requirements in formal epistemology, I will argue for the thesis in two ways from the premise that *akrasia* is irrational. I will then show that the Right Reasons position on peer disagreement follows from the Fixed Point Thesis and defend that position from arguments against it.

1

I first became interested in the Fixed Point Thesis by thinking about logical omniscience requirements in formal theories of rationality. The best-known such requirement comes from Bayesian epistemology, which takes Kolmogorov's probability axioms to represent rational requirements on an agent's degrees of belief. One of those axioms (usually called Normality) assigns a value of 1 to every logical truth. In Bayesian epistemology this entails that any agent who assigns less-than-certainty to a logical truth makes a rational error. Logical omniscience in some form is also a requirement of such formal epistemologies as ranking theory and AGM theory.

Logical omniscience requirements provoke four major objections:

- There are infinitely many logical truths. An agent can't entertain infinitely many propositions, much less assign certainty to all of them. (Call this the Cognitive Capacity objection.)
- Some logical truths are so complex or obscure that it isn't a rational failure to fail to recognize them as such and assign the required certainty. (Call this the Cognitive Reach objection.)

---

word we want to use to describe that notion). That is the notion I intend to discuss in this essay, and will use the word "rationality" to designate. Later on I'll consider whether the Fixed Point Thesis would be true if framed in terms of other normative notions (justification, reasons, etc.).

[3]The main exception I can think of is Descartes' *cogito* argument, which (with some major reinterpretation of Descartes' original presentation) could be read as an argument that it's irrational for an agent to believe she doesn't exist.

- Rational requirements are requirements of consistency *among* attitudes towards propositions. They do not dictate particular attitudes towards *single* propositions, as logical omniscience suggests.[4]
- Logical truths play no different role in the theory of rationality than any other truths, and rationality does not require certainty in all truths. Garber (1983, p. 105) writes, "Asymmetry in the treatment of logical and empirical knowledge is, on the face of it, absurd. It should be no more *irrational* to fail to know the least prime number greater than one million than it is to fail to know the number of volumes in the Library of Congress."

The last two objections seem the most challenging to me. (In fact, much of this essay can be read as a response to these two objections when applied to attitudes towards the rational requirements instead of attitudes towards logical truths.) The first two objections are rather straightforwardly met. For Cognitive Capacity, one need only understand the relevant logical omniscience requirements as taking the form "*If* one takes an attitude towards a logical truth, *then* one should assign certainty to it." Logical omniscience then does not require that attitudes be taken towards any particular propositions (or every member of any infinite set of propositions) at all.

To respond to the Cognitive Reach concern, we can restrict logical omniscience so that it requires certainty only in logical truths that are sufficiently obvious or accessible to the agent. Notice that even if we respond to the Cognitive Capacity and Cognitive Reach objections as I've just suggested, the other two objections remain: Why should a theory of rationality be in the business of dictating particular attitudes towards particular propositions (that is, if attitudes towards those propositions are taken at all), and why should the class of logical truths (even when restricted to the class of obvious logical truths) have a special status in the theory of rationality? Of course, filling out a plausible obviousness/accessibility restriction on the logical omniscience requirement is no trivial matter. One has to specify what one means by "obviousness," "accessibility," or whatever, and then one has to give some account of which truths meet that criterion in which situations. But since it was the objector who introduced the notion of obviousness or accessibility as a constraint on what can be rationally required, the objector is just as much on the hook for an account of this notion as the defender of logical omniscience.

Various writers have tried to flesh out reasonable boundaries on cognitive reach ((Cherniak 1986), for instance), and formal theories of rationality can be amended so as not to require full logical omniscience. Garber (1983) and Eells (1985), for example, constructed Bayesian formalisms that allow agents to be less-than-certain of first-order logical truths. Yet it is an underappreciated fact that while one can weaken the logical omniscience requirements of a formal epistemological theory, one cannot eliminate them entirely. The theories of Garber and Eells, for example, still require agents to be omniscient about the truths of sentential logic.[5]

---

[4] I am grateful to Alan Hájek for first bringing this objection to my attention; I have heard it from a number of people since then. There are echoes here of Hegel's famous complaint against Kant's categorical imperative that one cannot generate substantive restrictions from purely formal constraints.

[5] Gaifman (2004) takes a different approach to limiting Bayesian logical omniscience, on which the dividing line between what's required and what's not is not so tidy as sentential versus first-order. Still, there remains a class of logical truths to which a given agent is required to assign certainty.

Those wary of formal theorizing might suspect that this inability to entirely rid ourselves of logical omniscience is an artifact of formalization. But one can obtain logical omniscience requirements from informal epistemic principles as well. Consider:

>**Confidence:** Rationality requires an agent to be at least as confident of a proposition y as she is of any proposition x that entails it.

This principle is appealing if one thinks of an agent as spreading her confidence over possible worlds; since every world in proposition x is also contained in proposition y, the agent should be at least as confident of y as x. But even without possible worlds, Confidence is bolstered by the thought that it would be exceedingly odd for an agent to be more confident that the Yankees will *win* this year's World Series than she is that the Yankees will *participate* in that Series.

Given classical logic (which I will assume for the rest of this essay) it follows immediately from Confidence that rationality requires an agent to be equally confident of all logical truths and at least as confident of a logical truth as she is of any other proposition. This is because any proposition entails a logical truth and logical truths entail each other. One can add caveats to Confidence to address Cognitive Capacity and Reach concerns, but one will still have the result that if an agent assigns any attitude to a sufficiently obvious logical truth her confidence in it must be maximal.[6]

Confidence also reveals a connection between logical omniscience and the main topic of this essay: agents' attitudes towards truths about rational requirements. If one adopts Kant's understanding of logical truths, that connection is immediate. Kant (1974) took logical truths to be expressions of the rules of rational inference. So for Kant, a requirement that one be maximally confident in logical truths just is a requirement that one remain confident in the facts about rational requirements.

But Kant's account of logical truths is not common coin. So consider instead the Deduction Theorem of classical logic. The Deduction Theorem says that x entails y just in case $x \supset y$ is a logical truth. We can therefore rewrite Confidence as:

>**Confidence (reformulated):** If $x \supset y$ is a logical truth, rationality requires an agent to be at least as confident of y as x.

Take an agent who understands this principle (and believes it). Instances of the principle are specific rational requirements of the form "I should be at least as confident of y as x" for various xs and ys. In order for the agent to figure out what those specific rational requirements are, the agent has to determine which propositions of the form $x \supset y$ are logical truths. But that's a set of determinations logical omniscience requires the agent to get right. The specific *content* of Confidence's rational requirements is determined by the logical truth of material conditionals; logical omniscience (a consequence of Confidence itself) requires agents to get that content right. Logical truths (or at least logically true material conditionals) have a special status in the theory of rationality because they determine part of the content of that theory.

This line of thought is a bit fast and loose, but hopefully it brings the connection between logical omniscience and the Fixed Point Thesis into better focus. We can

---

[6]Notice that this argument makes no assumption that the agent's levels of confidence are numerical, or numerically representable.

sharpen that focus even more by observing what happens to formal epistemological theories when their logical omniscience requirement is weakened.

Formal theories don't require logical omniscience because formal theorists like the requirement; logical omniscience is a side effect of systems capturing the rational requirements theorists are actually after. Take the Bayesian case. Bayesian systems are designed to capture relations of rational consistency among attitudes and relations of confirmation among propositions. As I already mentioned, one can construct a Bayesian system that does not fault agents for failing to be certain of first-order logical truths. For example, one can have a Bayesian model in which an agent assigns credence less-than-1 to $(\forall x)Mx \supset Ms$. Applied to a sample consisting entirely of men, this model allows an agent to be less-than-certain that if all men are mortal then the man Socrates is as well. But in that model it may also be the case that Ms no longer confirms $(\forall x)Mx$, which is one of the basic confirmation relations one builds a Bayesian system to capture.[7] Similarly, in the imagined model the agent may no longer assign at least as great a credence to Ms as $(\forall x)Mx$; it will be possible for the agent to be less confident that the man Socrates is mortal than she is that all men are mortal.[8]

This is but one example of a second underappreciated fact: You cannot give up logical omniscience requirements without also giving up rational requirements on consistency and inference.[9] What is often viewed as a bug of formal epistemologies is necessary for their best features. This second underappreciated fact explains the first; if one removed all the logical omniscience requirements from a formal theory, that theory would no longer be able to place any constraints on consistency and inference, and so would be vitiated entirely.

In general, a rational requirement on consistency or inference often stands or falls with a requirement on attitudes towards a particular proposition.[10] (I call such a proposition the "dual" of the requirement on consistency or inference.[11]) Stepping away from credences for a bit, let's take the binary-belief example of an agent Jane whose evidence includes the proposition $p\&q$ but who infers from that evidence $\sim q$. It certainly looks like Jane's overall belief state is rationally flawed—given Jane's

---

[7]Here's how that works: Suppose that, following Garber, our model assigns credences over a formal language with an atomic sentence A representing $(\forall x)Mx$ and an atomic sentence S representing Ms. If our model has a basic Regularity requirement and we stipulate that $P(A \supset S) = 1$, we get the result that $P(S \mid A) > P(S \mid \sim A)$, so S confirms A. But if $P(A \supset S)$ is allowed to be less than 1, this result is no longer guaranteed.

[8]Taking the Garberian model from note 7, if $P(A \supset S) = 1 - c$ then $P(A)$ can exceed $P(S)$ by as much as c.

[9]As Max Cresswell has been arguing for decades (see, for example, (Cresswell 1975)), a version of this problem besets theories that model logical non-omniscience using logically impossible worlds. Such theories cannot make good sense of logical connectives—if we can have a world in which p and q are both true but $p\&q$ is not, what exactly does "&" *mean*?—and so lose the ability to represent the very sentences they were meant to model reasoning about. (For more recent work on the difficulties of using impossible worlds to model logical non-omniscience, see (Bjerring ms).)

[10](Balcerak Jackson and Balcerak Jackson ms) offers another nice example of this phenomenon. If an agent can rationally infer y from x, she can complete a conditional proof demonstrating $x \supset y$. Going in the other direction, if the agent rationally believes $x \supset y$ a quick logical move makes it rational to infer y from x. So the rational permission to infer y from x stands or falls with a rational permission to believe the proposition $x \supset y$.

[11]Note that the duality relation need not be one-to-one: a given rational requirement may have multiple dual propositions, and a given proposition may serve as a dual for multiple rational requirements.

evidence she seems to be required to believe q (at least if she takes any attitude towards q at all). If we see Jane make this inference (Jane is thinking aloud and says, "Hmm... I know that $p \& q$, so from that I conclude $\sim q$"), we will tell her she has made a rational error. Now imagine Jane responds by saying, "I understand very well the inference I just made; I happen to believe that rationality forbids me from inferring q from $p \& q$—in fact, I believe that I am rationally required on such evidence to believe $\sim q$."

When we first observed Jane's rationally mistaken inference, we might have supposed that it was some sort of performance error, a reasoning step Jane made (or mis-made) unthinkingly and without consciously applying any principle. But now we discover that Jane's inference error is accompanied by a belief about the requirements of rationality that seems (to her at least) to license that error. At this point I think it's very natural to read Jane as having a mistaken belief about the requirements of rationality, and to apportion that belief just as much rational opprobium as we initially gave to the inference error.

Now what if Jane goes on to cite (what she considers) evidence for her position about what rationality requires of her? Perhaps some respected authority on rational inference told her she was required to infer $\sim q$ from $p \& q$, or perhaps she was raised in a culture in which everyone made that sort of inference and she's never seen any contrasting behavior. At that point, I think three readings of Jane's situation are available:

(1) Jane's overall belief state is rationally flawed. Even given her total evidence (including any testimony, cultural influence, etc. concerning what rationality requires), the only attitude towards q rationally permissible for her is belief, and she is rationally required not to believe that belief in q is forbidden for her/belief in $\sim q$ is required. [call this the "bottom-up" reading]

(2) Jane's overall belief state is rationally flawed. Even given her total evidence the only attitude towards q rationally permissible for her is belief. Nevertheless, if her testimonial/cultural/etc. evidence about the requirements of rationality meets certain criteria, her (false) belief about what's required on her evidence is rationally permissible. [the "mismatch" reading]

(3) If her evidence about the requirements of rationality meets certain criteria, Jane's belief state is not rationally flawed at all. [the "top-down" reading]

How do these three readings interrelate—and how could the last one be remotely plausible? The bottom-up and mismatch readings share a basic assumption that whenever one's evidence includes a proposition of the form $x \& y$, the only rationally permissible attitude towards y is belief.[12] The bottom-up reading holds that Jane has made a rational mistake in getting this fact about rationality wrong. (This is the reading that follows from the Fixed-Point Thesis, and is the reading I will eventually defend.) The mismatch reading holds that circumstances could make it rationally permissible for Jane to have this mistaken belief about the requirements of rationality. For our purposes we need not precisely specify what criteria those

---

[12]Could that assumption hold up in the face of Cognitive Reach-style concerns? I hope conjunction elimination is a sufficiently blatant logical maneuver to clear whatever bar for obviousness Cognitive Reach objectors may set. If not, we can either switch the example to an even more obvious/accessible inference (if there is one!) that clears the bar, or we can add to the story that the logical relations between $p \& q$ and q have been made salient to Jane in all sorts of ways.

circumstances have to meet; they simply have to provide strong enough support for Jane's belief about what's required to make that belief rationally permissible. Even then, the mismatch reading holds that these circumstances don't make Jane's *inference* to $\sim q$ rationally permissible; they just make her belief about what's required okay.

The top-down reading agrees with the mismatch reading that Jane could receive evidence making it rationally permissible to believe that an inference from $p \,\&\, q$ to $\sim q$ is required, but according to the top-down reading this permission "trickles down" to rationally allow Jane to make that inference. One might motivate this reading by thinking about the fact that rational requirements are consistency requirements and concluding that as long as an agent's inferences are consistent with her beliefs about the rationality of those inferences, she is doing okay by rationality's lights. On this reading Jane's belief state is rationally permissible in all respects, because inferring $\sim q$ from $p \,\&\, q$ is a correct application of the rational principles in which she believes.

Notice that the top-down reading of Jane's case denies the other readings' assumption that any body of evidence containing $p \,\&\, q$ must render a belief in $\sim q$ rationally impermissible. The idea is that with the right kind of evidence about what's rationally permissible, a body of total evidence containing $p \,\&\, q$ might make belief in $\sim q$ rationally okay. At this point we might try finding a different example in which the link between a proposition and a conclusion is so strong that no body of evidence containing that proposition could permit disbelief in that conclusion. If the defender of a top-down reading of Jane's case concedes the existence of such examples, he will have to give either a bottom-up or a mismatch reading about those. Similarly, the bottom-up reader of Jane's case might allow other cases in which an agent's evidence about what's rational is given more sway.

But I think it's more philosophically interesting to take these two readings as instances of general positions. On a committed top-down view, no matter what components one's evidence contains that by themselves would rationally require a particular attitude towards a particular conclusion, there can be further components of that evidence that make it rational to believe that the contrary conclusion is required and thereby make it rational to draw that contrary conclusion. On this position, no evidence is ultimately indefeasible. Meanwhile, a committed bottom-up view holds that when a body of evidence requires a particular attitude towards a conclusion, no additional evidence about the rational requirements in that situation can make it rationally permissible to adopt a different attitude.[13]

The bottom-up and top-down readings of Jane's case do have one thing in common. They both embrace the claim I made before that rational requirements on

---

[13]The bottom-up view will admit that in many situations in which a body of evidence requires a particular attitude towards a particular conclusion, *some* kinds of additional evidence can change the permitted attitudes towards that conclusion. For example, when the original evidence doesn't entail the conclusion the addition of run-of-the-mill defeaters is possible. Also, depending on one's conception of evidence, additional evidence might cause an agent to call into question whether elements of her original evidence were true. (Perhaps Jane could learn something further that would cause her to doubt that $p \,\&\, q$.) The dispute between the bottom-up and top-down views is strictly about additional evidence concerning the rational requirements in the agent's situation.

consistency or inference stand or fall with requirements on attitudes towards particular propositions.[14] The proposition Jane believes about the permissibility of her inference is the dual of that inference itself. On the bottom-up reading, both the inference and her belief about the inference are rationally impermissible. On the top-down reading there are circumstances in which Jane's belief about the inference is rationally permissible, but in those circumstances the inference is permissible as well. Only the mismatch reading suggests that Jane could be permitted to believe that an inference is forbidden when that inference is in fact required.

I'll begin the next section by articulating an anti-*akrasia* principle that renders the mismatch reading false. I'll then argue from this principle to a limited version of the Fixed Point Thesis. Section 3 will complete the argument for the Fixed Point Thesis and eliminate the top-down view.

<div align="center">2</div>

2.1. Before I can argue for the Fixed Point Thesis, I need to define some terms and clarify the kinds of normative claims I will be making. We will be discussing both an agent's doxastic attitudes (for the time being just belief, disbelief, or suspension of judgment—we'll get back to credences later) and her intentions. I will group both doxastic attitudes and intentions under the general term "attitudes." Because some of the rational rules we'll be discussing impugn combinations of attitudes without necessarily indicting individual attitudes within those combinations, I will not be evaluating attitudes in isolation. Instead I will examine rational evaluations of an agent's "overall state," which includes all the attitudes she assigns at a given time.

Evaluations of theoretical rationality concern only the doxastic attitudes in an agent's overall state. Evaluations of practical rationality may involve both beliefs and intentions. For an example, there might be a (wide-scope) requirement of instrumental rationality that negatively evaluates any overall state including an intention to ϕ, a belief that ψ-ing is necessary for ϕ-ing, and an intention not to ψ.[15]

Having discussed what rational requirements are requirements *on*, we now need to discuss what rational requirements *supervene* on. Different positions are available here. For example, an evidentialist holds that rational requirements on doxastic attitudes supervene on an agent's evidence. We might then think that the rational requirements on an agent's overall state (both beliefs and intentions) supervene on her evidence and her desires. But we might think instead that requirements on intentions supervene on reasons, not desires. Or if we want to deny evidentialism, we might suggest that requirements on beliefs supervene not just on an agent's evidence but also on, say, her past beliefs.[16]

---

[14]To be clear, I am *not* suggesting that every time an agent makes an inference error she also has a mistaken belief about the requirements of rationality; plenty of poor inferrers have never even thought about the requirements of rationality. However we can *generate* plenty of cases in which the agent has explicit higher-level views, and then argue that in such cases the requirements at different levels match.

[15]One might think that some requirements of practical rationality involve not just on an agent's intentions but also on her actions. In that case one would have to include actions in an agent's overall state along with her attitudes. For simplicity's sake I'm going to focus just on rational evaluations involving beliefs and intentions.

[16]Various versions of conservatism and coherentism in epistemology take this position.

To remain neutral on these points I will assume only that whatever the true theory of rationality is, it specifies certain aspects of an agent's circumstances as relevant to determining the rational requirements on her overall state. Taken together, these relevant aspects comprise what I'll call the agent's "situation." An agent's situation at a given time probably includes features of her condition at that time, but it might also include facts about her past or other kinds of facts.

Given an agent's current situation and overall state, we can evaluate her state against her situation to see if the state contains any rational flaws. That is, we can ask whether from a rational point of view there is anything negative to say about the agent's adopting that overall state in that situation. This is meant to be an *evaluative* excercise that does not immediately lead to any prescriptions—I am not suggesting a rational rule that agents ought only adopt rationally flawless states. In the conclusion I will assess the significance of such evaluations of rational flawlessness.

But in the meantime we have a more pressing problem. I want to be able to say that in a given situation some particular overall states are rationally without flaw, and even sometimes that a particular overall state is the only flawless state available in a situation. But English has no concise, elegant way to say things like that, especially when we want to put them in verb phrases and the like. So I am going to repurpose a terminology already to hand for describing states that satisfy all the principles of a kind and states that uniquely satisfy such principles: I will describe an overall state with no rational flaws as "rationally permissible." A state that is not rationally permissible will be "rationally forbidden." And if only one overall state is flawless in a given situation, I will call that state "rationally required."[17]

I will also apply this terminology to individual attitudes. If an agent's current situation permits at least one overall state containing a particular attitude, I will say that that attitude is rationally permissible in that situation. If *no* permitted states contain a particular attitude, I will say that attitude is rationally forbidden in the current situation. If *all* permitted states contain an attitude I will say that attitde is rationally required. Notice, however, that while talking about attitudes this way is a convenient shorthand, it is a shorthand for evaluations of *entire* states; at no point am I actually evaluating attitudes in isolation.[18]

I realize that the "permitted" and "required" terminology I've repurposed here usually carries prescriptive connotations—we'll simply have to remind ourselves periodically that we are discussing a purely evaluative project. I also want to emphasize that I am evaluating states, not agents, and I certainly don't want to get into assignations of praise or blame. At the same time the states being evaluated are states of real agents, not states of mythical idealized agents. Even if you're convinced that a real agent could never achieve a rationally flawless set of attitudes, it can be interesting to consider what kinds of rational flaws may arise in an agent's attitude set. Finally, my rational evaluations are all-things-considered

---

[17]Situations that allow for no rationally flawless overall states are rational dilemmas.

[18]If we wanted, we could represent the requirements of rationality using a function. Given any situation S, $\mathcal{R}(S)$ would be the set of rationally flawless overall states in situation S. Each overall state, in turn, would be a set of doxastic attitudes and intentions. Notice that these overall states could be "partial," in the sense that they need not contain a doxastic attitude towards every proposition or an intention concerning every possible action. That would be consistent with my earlier response to the Cognitive Capacity objection.

evaluations. I will be asking whether, given an agent's current situation and taking into account *every* aspect of that situation pointing in whatever direction, it is all-things-considered rationally permissible for her to adopt a particular combination of attitudes.

Working with situations and overall states, we can characterize a variety of theses about rationality. There might, for instance, be a rational rule about perceptual evidence that if an agent's situation includes a perception that x, all the overall states rationally permissible for her include a belief that x. Such a rule relates an agent's beliefs to her evidence; other rational rules might embody consistency requirements strictly among an agent's beliefs.[19] Perhaps no situation rationally permits an overall state containing logically contradictory beliefs, or perhaps there's an instrumental ϕ/ψ rationality requirement of the sort described earlier. On the other hand, there may be no general rules of rationality at all. But even a particularist will admit that there are rational requirements on agents' overall states in particular situations; he just won't think any general, systematic characterizations of such constraints are available.

Using this terminology, the Fixed Point Thesis becomes:

> **Fixed Point Thesis (reformulated):** No situation rationally permits an *a priori* false belief about which states are rationally permitted in which situations.[20]

There are still some imprecisions in this statement (for instance, the relevant notion of "about"ness needs cleaning up), but I hope the general thrust is clear.[21]

I will soon be offering two arguments for the Fixed Point Thesis. Both arguments make the following assumption:

> **Anti-Akrasia:** No situation rationally permits any overall state containing both an attitude A and the belief that A is rationally forbidden in one's current situation.

Anti-Akrasia says that any akratic overall state is rationally flawed in some respect. It applies both to cases in which an agent has an intention A while believing that intention is rationally forbidden, and cases in which the agent has a belief A that she takes to be rationally forbidden in her situation. Anti-Akrasia does not come down on whether the rational flaw is in the agent's intention (say), in her belief about the intention's rational status, or somehow in the combination of the two. It simply says that if an agent has such a combination in her overall state, that state is rationally flawed. Thus Anti-Akrasia is a wide-scope norm; it does *not* say that whenever an agent believes A is forbidden in her situation that agent is in fact forbidden to assign A.[22]

---

[19]The distinction is simply meant to be illustrative; I am not making any substantive assumption going forward that an agent's evidence is not a subset of her beliefs.

[20]In terms of the function from note 18, this becomes: For no S does $\mathcal{R}(S)$ include an overall state containing an *a priori* false belief about the values of $\mathcal{R}(S)$.

[21]Notice that even if the situation determining rational requirements on an agent can include empirical facts not accessible to the agent (such as facts about her beliefs in the past), there will still be *a priori* truths about which situations generate which rational requirements. They will take the form "*if* the empirical facts are such-and-such, *then* rationality requires so-and-so."

[22]Arpaly (2000) argues (*contra* Michael Smith and others) that in some cases in which an agent has made an irrational mistake about which attitude rationality requires, it can still be rationally better for him to adopt the rationally-required attitude than the one he thinks is required. In this case Anti-Akrasia indicates that if the agent adopts the rationally-required attitude then his

The rationality of practical *akrasia* has been discussed for centuries (if not millenia), and I take it the overwhelming current consensus endorses Anti-Akrasia for the practical case. Discussions of the theoretical case (in which A is a belief) tend to be more recent and rare. Feldman (2005) discusses a requirement on beliefs he calls "Respect Your Evidence," and for anyone who doubts Anti-Akrasia in the belief case it is well worth reading Feldman's defense.[23] (Requirements like Respect Your Evidence are also discussed in (Adler 2002), (Bergmann 2005), (Gibbons 2006), and (Christensen 2010).) Among other things, Feldman points out that an agent who violated Anti-Akrasia for beliefs could after a quick logical step find herself with a Moore-paradoxical belief of the form "x, but it's irrational for me to believe x."[24]

Nevertheless, Brian Weatherson argues against a version of Anti-Akrasia.[25] He begins with the following example (quoted from (Weatherson ms, p. 12)):

> **Kantians:** Frances believes that lying is morally permissible when the purpose of the lie is to prevent the recipient of the lie performing a seriously immoral act. In fact she's correct; if you know that someone will commit a seriously immoral act unless you lie, then you should lie. Unfortunately, this belief of Frances's is subsequently undermined when she goes to university and takes courses from brilliant Kantian professors. Frances knows that the reasons her professors advance for the immorality of lying are much stronger than the reasons she can advance for her earlier moral beliefs. After one particularly brilliant lecture, Frances is at home when a man comes to the door with a large axe. He says he is looking for Frances's flatmate, and plans to kill him, and asks Frances where her flatmate is. If Frances says, "He's at the police station across the road", the axeman will head over there, and be arrested. But that would be a lie. Saying anything else, or saying nothing at all, will put her flatmate at great risk, since in fact he's hiding under a desk six feet behind Frances. What should she do?

Weatherson responds to this example as follows:

> That's an easy one! The text says that if someone will commit a seriously immoral act unless you lie, you should lie. So Frances should lie. The trickier question is what she should believe. I think she should believe that she'd be doing the wrong thing if she lies. After all, she has excellent evidence for that, from the testimony of ethical experts, and she doesn't have compelling defeaters for that testimony. So she should do something that she believes, and should believe, is wrong. . . .

---

overall state is rationally flawed. That is consistent with Arpaly's position, since she has granted that the agent's belief in this case about what's rationally required already creates a rational flaw in his overall state. (Arpaly explicitly concedes the presence of that rational flaw at (Arpaly 2000, p. 491).)

[23]Since Feldman is an evidentialist, he takes an agent's situation (for belief-evaluation purposes) to consist solely of that agent's evidence. His principle also concerns justification rather than rationality.

[24]See (Smithies 2011) for further discussion of such paradoxical statements.

[25]Weatherson's ultimately wants to deny a version of Anti-Akrasia for the theoretical case. But he gets there by first arguing against a version of Anti-Akrasia for the practical case, and then drawing an analogy between the practical and the theoretical.

> For her to be as she should, she must do something she believes is wrong. That is, she should do something even though she should believe that she should not do it. So I conclude that it is possible that sometimes what we should do is the opposite of what we should believe we should do.

There are a number of differences between our Anti-Akrasia and the principle Weatherson is attacking. First, we are considering *intentions* while Weatherson is considering what *actions* Frances should perform. So let's suppose Weatherson also takes this example to establish that sometimes what intention we should form is the opposite of what intention we should believe we should form. Second, Weatherson is considering what attitudes Frances *should* have, while we're considering what combinations of attitudes would be *rationally flawed* for Frances to have. Can Weatherson's Frances example be used to argue against our Anti-Akrasia principle, concerning rationally-flawed overall states?

When we try to use Frances to build such an argument, the case's description immediately becomes tendentious. Transposed into rationality-talk, the second sentence of the Kantians description would become, "If you know that someone will commit a seriously immoral act unless you lie, you are rationally required to lie." This blanket statement rules out the possibility that what an agent is rationally required to do in the face of someone about to commit a seriously immoral act might depend on what evidence that agent has about the truth of various ethical theories. The top-down theorist from section 1 will insist that if Frances has enough reason to believe that Kantian ethics is true, then in fact Frances is rationally forbidden from intending to lie to the axeman at the door. (And thus is not required to behave in a way she believes is rationally forbidden.) In other words, the top-down theorist will deny that the case Weatherson has described (transposed into rationality-talk) is possible. Going in the other direction, a bottom-up theorist might refuse to concede Weatherson's claim that Frances "doesn't have compelling defeaters for" the testimony of her professors. If rationality truly requires intending to lie to the axeman, whatever reasons make that the case will also count as defeaters for the professors' claims.

These responses bring out something odd about Weatherson's reading of the Kantian case. Imagine you are talking to Frances, and she is wondering whether she is rationally required to believe what her professor says. To convince her that she is, there are various considerations you might cite—the professor knows a lot about ethics, he has thought about the case deeply and at great length, he has been correct on many occasions before, etc.—and presumably Frances would find some of these considerations convincing.[26] Now suppose that instead of wondering whether she is required to *believe* what her professor says, Frances comes to you and asks whether she is required to *intend* as her professor prescribes. It seems like the points you made in the other case—the professor knows a lot about how one ought to behave, he has thought about her kind of situation deeply and at great length, he has prescribed the correct behavior on many occasions before, etc.—apply equally

---

[26]I am not taking a position here on whether testimony is in some sense a "fundamental" source of justification. Even if testimonial justification is fundamental, one can still adduce considerations to an audience that will make accepting testimony seem appealing. Fundamentalism about testimonial justification is not meant to choke off all discussion of whether believing testimony is epistemically desirable.

well here. That is, any consideration in favor of believing what the professor says is also a consideration in favor of behaving as the professor suggests, and *vice versa*. There is a sort of parity thesis here:

> **Testimonial Parity:** When testimony describes a particular course as rationally required (or forbidden), one is rationally required to believe the testimony just in case one is rationally required (/forbidden) to do what's described.

Notice that I have not argued for Testimonial Parity on the basis of Anti-Akrasia. Testimonial Parity follows not from considerations of coherence among an agent's beliefs and actions (or intentions), but instead from thinking about what can be said in favor of the belief and what can be said in favor of the action. If Testimonial Parity is correct, Weatherson cannot just stipulate in the Kantian case what Frances is required to do, then go on to describe what her professor says and claim that she is bound by that as well. The professor's testimony may give Frances reasons to behave differently than she would otherwise, or the moral considerations involved may give Frances reason not to believe the testimony. So I don't think Kantians provides a convincing counterexample to Anti-Akrasia.[27]

There is another kind of case in which what an agent should do might diverge from what she should believe she should do. My argument for Testimonial Parity assumed that anything that's a reason to believe a piece of testimony can also be a reason to obey it, and *vice versa*. But we can have cases in which certain reasons bear on behavior but not on belief. To see this possibility, consider Bernard Williams's famous example (Williams 1981) of the agent faced with a glass full of petrol who thinks it's filled with gin. For Williams, what an agent has reason to do is determined in part by what that agent would be disposed to do if she were *fully* informed. Thus the fact that the glass contains petrol gives the agent reason not to drink what's in it. But this fact does not give the agent reason to believe that the glass contains petrol, and so does not give the agent any reason to believe she shouldn't drink its contents. For Williams, any true fact in the universe may provide an agent with reason to behave in particular ways. Yet we tend to think that an agent's reasons to believe include only *cognitively local* facts. A position on which an agent has reason to believe only what she would believe were she *fully* informed makes all falsehoods impermissible to believe (and makes all-things-considered misleading evidence impossible in every case).

If we accept this difference between the dependence bases of practical and theoretical reasons, it's reasonable to hold that an agent can have most reason to act (or intend) in one direction while having most reason to believe she should act in another. What the agent has reason to *believe* about whether to drink the liquid in front of her is determined by her cognitively local information; what she has reason to *do* is determined by an unrestricted set of facts.[28] And if we think that what

---

[27]Note that Kantians could be a rational dilemma—a situation in which no overall state is rationally permitted. In that case Kantians would not be a counterexample to Anti-Akrasia because it would not constitute a situation in which an overall state is permitted containing both an attitude A and the belief that that attitude is forbidden. We will return to rational dilemmas in section 4.

[28]It's interesting to consider whether one could get a similar split between what an agent has reason to believe and what she has reason to believe *about* what she has reason to believe. If there is some boundary specifying how cognitively local a fact has to be for it to count as a reason for belief, then the dependency bases for an agent's reasons for first-order beliefs and higher-order

an agent *should* do or believe supervenes on what she has *most reason* to do or believe, we might be able to generate cases in which an agent should do one thing while believing that she should do another.

Yet here we return to potential distinctions between what an agent should do, what she has most reason to do, and what she is rationally required to do.[29] It's implausible that in Williams's example the agent is rationally required to believe the glass contains gin but rationally forbidden to drink what's in it.[30] What one is rationally required to do or believe depends only on what's cognitively local—this is what drove the accessibility concerns of the Cognitive Reach objection. As long as the normative notion put to work in Anti-Akrasia is rational requirement, Williams-style cases don't generate counterexamples to the principle. So I will continue to side with the strong, generally-shared intuition that an akratic agent is (in John Broome's (1999) evocative phrase) "not entirely as she ought to be" from a rational point of view.

But this discussion reveals something important about the Fixed-Point Thesis and the arguments for it I will soon provide. While I have framed the Fixed Point Thesis in terms of rational requirements, one might wonder whether it applies equally to other normative notions. (Could one be *justified* in a mistake about *justification*? Could one have *most reason* for a false belief about what *reasons* there are?) I am going to argue for the Fixed Point Thesis on the basis of Anti-Akrasia, which is about rational requirements. As we've just seen, Anti-Akrasia might be less plausible for other normative notions: for the notion of "reasons," Williams-style cases might be available; and for various kinds of "oughts," Weatherson's Kantians might be convincing. For any normative notion for which an analogue of Anti-Akrasia holds, I believe I could run my arguments for a version of the Fixed-Point Thesis featuring that normative notion. But for normative notions for which a version Anti-Akrasia is not plausible, I do not know if a Fixed-Point analogue would be supportable.

2.2. My first argument will actually be for a restricted version of the Fixed Point Thesis:

> **Special Case Thesis:** There do not exist an attitude A and a situation such that:
> - A is rationally required in the situation, yet
> - it is rationally permissible in that situation to believe that A is rationally forbidden.

As a special case of the Fixed Point Thesis (concerning a particular *kind* of mistake about the rational requirements that an agent could make) the Special Case Thesis is logically weaker than the Fixed Point Thesis. However, I think anyone inclined to deny the Fixed Point Thesis will be inclined to deny its application to this special

---

beliefs would be the same. In that case, it seems difficult to generate a Williams-style case in which an agent has reason to believe one thing but reason to believe that she has reason to believe another, because we don't have the excuse that the former can draw on sets of facts not available to the latter. In the end, this might make it even more difficult to deny versions of Anti-Akrasia for the theoretical case (in which A is a doxastic attitude) than for the practical case (in which A is an intention).

[29]Not to mention the distinction between an agent's "subjective" and "objective" reasons. (See (Schroeder 2008) for a careful examination of the intersection of that distinction with the issues considered here.)

[30]Despite an odd suggestion to that effect at (Williams 1981, pp. 102–103).

case as well. ("What if the agent was raised to believe she's forbidden to assign A?!?")

My first argument for the Special Case Thesis is called No Way Out. The argument is a *reductio*. Begin by supposing (contrary to the Special Case Thesis) that we have a case in which an agent's situation rationally requires the attitude A yet also rationally permits an overall state containing a belief that A is rationally forbidden. Now consider that permitted overall state, and ask whether A appears in it or not. If the permitted overall state does not contain A, we have a contradiction with our supposition that the agent's situation requires A. (That supposition simply says that every overall state rationally permissible in the situation contains A.) So now suppose that the permitted overall state includes A. Then the state includes both A and a belief that A is forbidden in the current situation. By Anti-Akrasia this state is not rationally permissible, contrary to supposition once more. This completes our *reductio*. Anti-Akrasia entails the Special Case Thesis.[31]

It's surprising that the Special Case Thesis is so straightforwardly derivable from Anti-Akrasia. Part of the surprise comes from deriving something controversial (if not downright counterintuitive) from something almost every philosopher believes. But I think another part of the surprise comes from deriving a *substantive* conclusion from a *structural* premise. Here I am borrowing terminology from Scanlon (2003), though not using it exactly as he does.[32] Structural constraints concern the way an agent's attitudes hang together, while substantive constraints explain which particular attitudes an agent's situation requires of her. In epistemology, structural norms of coherence and consistency among an agent's beliefs are often contrasted with substantive norms about how her beliefs should be driven by her evidence.

If one accepts this division, Anti-Akrasia certainly looks like a structural rationality claim. The Special Case Thesis, meanwhile, says that when a particular fact is *true* in an agent's situation she is forbidden from disbelieving it in a certain way. The No Way Out argument moves from a premise about the general consistency of an agent's attitudes to a conclusion about what the specific content of those attitudes must be.

---

[31]One might grant the truth of the Special Case Thesis but hold that it is only trivially true, because there is no attitude A and situation for which the situation rationally *requires* A. The thought would be that it is always rationally permissible for an agent to fail to take a given attitude A. This is incredibly implausible for As that are intentions. If any actions are ever rationally required by moral principles, or instrumental principles, or pragmatic principles, etc., then the Special Case Thesis is non-trivial. As for applying the Special Case Thesis to beliefs, I have already expressed sympathy for the idea that an agent can't be rationally required to believe a proposition that, say, she has never entertained. But keep in mind that an agent's situation includes *all* the facts on which requirements on her overall state supervene. So, for instance, an agent's situation may include the facts that she has entertained a particular proposition, thoroughly considered its connection to her evidence, and even realized that that proposition is entailed by her evidence. In that kind of situation we may agree that rationality *requires* her to form a belief in the given proposition.

[32]Scanlon distinguishes structural normative claims from substantive normative claims. Scanlon works in terms of reasons, and has a particular view about how the structural claims are to be understood, so he distinguishes structural from substantive normative claims by saying that the former "involve claims about what a person must, if she is not irrational, treat as a reason, but they make no claims about whether this actually *is* a reason." (Scanlon 2003, p. 13, emphasis in original) There's also the issue that in his earlier writings (such as (Scanlon 1998)) Scanlon claimed only structural claims had to do with *rationality*, but by (Scanlon 2003) he had withdrawn that assertion.

That conclusion—the Special Case Thesis—may seem to run afoul of our earlier Cognitive Reach concerns. The thesis forbids believing that A is rationally forbidden whenever it's simply *true* that A is required; no mention is made of whether A's being required is sufficiently accessible or obvious to the agent. This makes Special Case seem like an externalist thesis, in epistemologists' sense of "externalist."[33] But that appearance is incorrect. Suppose you hold that in order for an attitude to be rationally required (or forbidden) of an agent in a situation, the relevant relation between the situation and that attitude must be sufficiently accessible or obvious to the agent. Under this view, whenever it's true that attitude A is required of an agent in a situation it's also true that A's relation to the situation is sufficiently accessible or obvious to the agent. So whenever the Special Case Thesis applies to an agent, that agent has sufficiently obvious and accessible materials available to determine that it applies. The moment an internalist grants that *any* attitudes are required, he's also granted that there are propositions about rationality agents are forbidden to believe.

No Way Out has no consequences for the dispute between internalists and externalists in epistemology. But it does have consequences for the notion of evidential support. I said earlier that the evaluations discussed in our arguments are all-things-considered appraisals of rational permissibility. Most people hold that if an agent's total evidence supports a particular conclusion, it is at least rationally permissible for her to believe that conclusion. Yet the Special Case Thesis says there is never a case in which an attitude A is rationally required but it is rationally permissible to believe that attitude is forbidden. This means an agent's total evidence can never all-things-considered support the conclusion that an attitude is forbidden when that attitude is in fact required. Put another way, a particular type of all-things-considered misleading evidence about rational requirements is impossible. The No Way Out argument moves from a premise about consistency requirements among an agent's attitudes (Anti-Akrasia) to a strong conclusion about what can be substantively supported by an agent's evidence.

The point is a bit delicate. The top-down theorist will remind us that an agent's evidence about what is rationally forbidden or required of her may affect what is indeed forbidden or required. The Special Case Thesis reveals that whatever effects an agent's higher-order evidence has on what's required in her situation, it will never turn out in the end that an attitude is required of the agent while her total evidence points all-things-considered towards that attitude's being forbidden. Of the three positions we discussed in Section 1, the mismatch position is the only one ruled out by Anti-Akrasia, as the Special Case Thesis makes clear.[34]

I should also be clear that the Special Case Thesis is not the full Fixed Point Thesis. No Way Out concerns cases in which an agent makes a mistake about what's required by *her own* situation, and in which the mistake is as extreme as possible—the agent takes an attitude that's *required* to be *forbidden*. To reach the

---

[33]The Cognitive Reach concern need not be exclusive to (epistemological) externalists. While accessibility is an internalist concern, externalists who reject accessibility as a necessary requirement for various positive epistemic evaluations may nevertheless hold that a relation must be sufficiently *obvious* to an agent for it to rationally require something of her.

[34]The Special Case Thesis also shows how Anti-Akrasia generates instances of the duality phenomenon discussed in Section 1. Once one accepts Anti-Akrasia one has the Special Case Thesis, which says that whenever a situation rationally requires a particular attitude, there is a proposition the agent in that situation is rationally forbidden to believe.

full Fixed Point Thesis, we would have to generalize the Special Case Thesis in two ways:

(1) to mistakes about what's rationally required by situations other than the agent's own current situation; and
(2) to mistakes besides believing something required is forbidden.

As an example of the second generalization, we would for example have to treat cases in which an attitude is rationally required for an agent, the agent has that attitude, but the agent thinks the attitude is merely rationally permissible in her situation.

My next argument will not only reinforce the Special Case Thesis, but also support the first generalization above. A strategy for establishing the second generalization will emerge when we turn to peer disagreement.

2.3. One strong source of resistance to the Fixed Point Thesis is the intuition that if an agent has the right kind of evidence—testimonial, cultural, etc.—that evidence can rationally permit her to mistakenly believe that a particular belief is forbidden. No Way Out combats the intuition that evidence might authorize false beliefs about the requirements of rationality by showing that an agent who formed such beliefs would be in a rationally untenable position. But that doesn't explain where the intuition goes wrong; it doesn't illuminate *why* evidence can't all-things-considered support such false beliefs. My next argument, the Self-Undermining argument, focuses on what the requirements of rationality themselves would have to be like for these false beliefs to be rationally permissible.

Suppose, for example, that the following were a rule of rationality:

> **Testimony:** If an agent's situation includes testimony that x, the agent is rationally permitted and required to believe that x.

By saying that the agent is both permitted and required to believe that x, I mean that the agent's situation permits at least one overall state and all permitted overall states contain a belief that x. The permission part is important, because I'm imagining an interlocutor who thinks that an agent's receiving testimony that x makes it acceptable to believe that x even if x is false or epistemically undesirable in some other respect. Of course Testimony is drastically oversimplified in other ways, and in any case testimony is not the only source from which an agent could receive evidence about what's rationally required. But after presenting the Self-Undermining argument I'll suggest that removing the simplifications in Testimony or focusing on another kind of evidence would leave my main point intact.[35]

The Self-Undermining argument shows by *reductio* that Testimony cannot express a true general rule of rationality. So begin by supposing Testimony is true, then suppose that an agent receives testimony that p:

> If an agent's situation includes testimony that x, the agent is rationally forbidden to believe that x.

By Testimony, the agent in this situation is permitted an overall state in which she believes that p. So suppose the agent is in that rationally permitted state. We can now complete the *reductio* in one of two ways.

---

[35]A particularist may reject Testimony because he rejects all general rational rules. But such a particularist will also reject Anti-Akrasia as a general rational rule, and so will have gotten off our argumentative bus at a much earlier stop.

First, we can invoke Testimonial Parity. By Testimony, the agent is required to believe p, and since p describes a particular course as rationally forbidden, Testimonial Parity says that the agent is required to believe p just in case she is forbidden to do what p describes. Since p says the agent is forbidden to believe p in her situation (since p was a deliverance of testimony), the agent is rationally forbidden to believe p. But then her overall state, which includes a belief in p, is rationally impermissible, and we have a contradiction.

Of course one might simply reject Testimonial Parity, in which case we can complete the Self-Undermining argument in a second way. Since the agent believes p, she believes that it's rationally impermissible to believe testimony. She learned p from testimony, so she believes that belief in p is rationally forbidden in her situation. But now her overall state includes both a belief in p and a belief that believing p is rationally forbidden. By Anti-Akrasia, the agent's state is rationally impermissible, and once again we have a contradiction. Our argument has demonstrated that Testimonial Parity and Anti-Akrasia each independently entail that Testimony is not a true rule of rationality.

A moment ago I admitted that Testimony is drastically oversimplified as a putative rational rule, and one might think that adding in more realistic complications would allow Testimony to avoid Self-Undermining. For example, an agent isn't required and permitted to believe just *any* testimony she hears; that testimony must come from a particular kind of source. Instead of investigating exactly what criteria a source must meet for its testimony to be rationally convincing, I'll just suppose that such criteria have been identified and call any source meeting them an "authority." The Testimony rule would then say that an agent is required and permitted to believe testimony from an authority. And the thought would be that when the agent in the Self-Undermining argument hears her source say p, she should stop viewing that source as an authority. (Anyone who says something as crazy as p certainly shouldn't be regarded as an authority!) The source's testimony therefore doesn't generate any rational requirements or permissions for the agent, the argument can't get going, and there is no problem for the (suitably modified) Testimony rule.

Whatever the criteria are for being an authority, they cannot render the Testimony norm vacuous. That is, it can't be that a source qualifies as an authority by virtue of his testimony's rationally requiring and permitting an agent to believe what it says. Usually a source qualifies as an authority by virtue of being reliable, having a track-record of speaking the truth, being trusted, or some such. Whatever those criteria are, we can stipulate that the source providing testimony that p in the Self-Undermining argument has met those criteria. Then the claim that the agent should stop treating her source like an authority the moment that source says p really becomes a flat denial of the Testimony rule (even restricted to testimony from authorities). The position is no longer that all testimony from an authority permits and requires belief; the position is that authorities should be believed unless they say things like p.

This point about the "authorities" restriction generalizes. Whatever restrictions we build into the Testimony rule, it will be possible to construct a case in which the agent receives a piece of testimony satisfying those restrictions that nevertheless contradicts the rule. That is, it will be possible unless those restrictions include

a *de facto* prohibition of just such testimony.[36] At that point, it's simpler just to modify the Testimony rule as follows:

> **Restricted Testimony:** If an agent's situation includes testimony that x, the agent is rationally permitted and required to believe that x—unless x contradicts this rule.

This rule performs exactly like Testimony in the everyday cases that lend Testimony intuitive plausibility. But the added restriction innoculates the rule to Self-Undermining; it stops that argument at its very first step, in which the agent's receiving testimony that p makes it permissible for her to believe p. p contradicts Restricted Testimony by virtue of providing an opposite rational judgment from Restricted Testimony on all xs received via testimony that don't contradict the rule.[37] Thus the restriction in Restricted Testimony keeps testimony that p from rationally permitting or requiring the agent to believe that p.

There's nothing special about Testimony as a rational rule here. We're going to want all rational rules to be similarly restricted to prevent Self-Undermining. For example, we might have the following:

> **Restricted Perceptual Warrant:** If an agent's situation includes a perception that x, the agent is rationally required to believe that x—unless x contradicts this rule.
>
> **Restricted Closure:** In any situation, any rationally permitted overall state containing beliefs that jointly entail x also contains a belief that x—unless x contradicts this rule.

The restriction may be unnecessary for some rules because it is vacuous. (It's hard to imagine a situation in which an agent *perceives* a proposition that directly contradicts a rational rule.) But even for those rules, it does no harm to have the restriction in there.

Previous authors have noted the need to restrict rational rules from undermining themselves. (See, for example, (Elga 2010).[38]) But what hasn't been noticed is

---

[36]One might think that Testimony could be fixed by adding in a "no defeaters" clause, which blocks an agent from being permitted and required to believe testimony for which she has a defeater. One could then avoid Self-Undermining by arguing that p is a defeater for the very testimony that conveys it. I suspect that any such argument would wind up invoking the sort of *de facto* prohibition just mentioned in the text. But even worse, it seems to me that a proposition can defeat an agent's rational support for a conclusion only if it is rationally permissible for the agent to believe that proposition. In order for p to act as a defeater it would have to be rationally permissible for the agent to believe p, but if p is permissible to believe our Self-Undermining *reductio* starts up again (at least in its second version, invoking Anti-Akrasia).

[37]If we read both p and Restricted Testimony as material conditionals universally quantified over a domain of possible cases, then as it stands there is no direct logical contradiction between them—both conditionals could be satisfied if neither antecedent is ever made true. But if we assume as part of our background that the domain of possible cases includes some instances of testimony that don't contradict the rule, then relative to that assumption p and Restricted Testimony contradict each other.

[38]Elga's discussion takes off from a real-life case. Having advocated a Split the Difference position on peer disagreement like the one we'll discuss in Section 3, Elga found that many of his peers disagreed with that position. It then seemed that by his own lights Elga should give up his staunch adherence to Split the Difference. Elga's response is to argue that Split the Difference requires being conciliatory about all propositions except that rational rule itself.

For a real-life case of testimonial self-undermining, consider this exchange from the ballroom scene in *The Muppet Show* (Season 1, Episode 11): "I find that most people don't believe what other people tell them." "I don't think that's true."

that restrictions on *self*-undermining don't solve the problem. Rational rules must include not only exceptions to avoid undermining themselves; they must also include exceptions to avoid undermining each other. To see why, suppose for *reductio* that the three restricted rules just described are true. Now consider an unfortunate agent who both perceives that she has hands and receives testimony of the disjunction that either p is true or she has no hands (where p is as before). By Restricted Testimony, there is a rationally permissible state in that agent's situation in which the agent believes "Either p or I have no hands." (Notice that this belief does not logically contradict Restricted Testimony, and so does not invoke that rule's restriction.) By Restricted Perceptual Warrant, that permitted overall state also includes a belief that the agent has hands (which clearly doesn't contradict the restricted perceptual warrant rule). By Restricted Closure, that permitted state also contains a belief that p (which, while it contradicts Restricted Testimony, does not contradict Restricted Closure). But p indicates that the agent is rationally forbidden to believe "Either p or I have no hands," and we can complete our argument as before by invoking either Testimonial Parity or Anti-Akrasia.

At no point in this argument does one of our restricted rational rules dictate that a belief is required or permitted which logically contradicts *that rule*. Instead we have constructed a loop in which no rule undermines itself but together the rules wind up undermining each other.[39] Clearly we could expand this kind of loop to bring in other rational rules if we liked. And the loop could be constructed even if we added various complications to our perceptual warrant and closure rules to make them independently more plausible. For example, clauses added to Restricted Closure in response to Cognitive Capacity and Cognitive Reach concerns could be accommodated by stipulating that our unfortunate agent entertains all the propositions in question and recognizes all the entailments involved.

The way to avoid such loops is to move not from Testimony to Restricted Testimony but instead to:

> **Properly Restricted Testimony:** If an agent's situation includes testimony that x, the agent is rationally permitted and required to believe x—unless x contradicts an *a priori* truth about what rationality requires.

and likewise for the other rational rules.

These proper restrictions on rational rules explain the points about evidence that puzzled us before. Rational rules tell us what various situations permit or require.[40] Rational rules concerning belief reveal what conclusions are *supported* by various bodies of evidence. In typical, run-of-the mill cases a body of evidence containing testimony all-things-considered supports the conclusions that testimony contains, as will be reflected in most applications of Properly Restricted Testimony. But an agent may receive testimony that contradicts an (*a priori*) truth about the rational rules. Generalizing from typical cases, we thought that even in those cases the evidence supports what the testimony conveys. And so we thought it could be rationally permissible—or even rationally required—to form beliefs at odds with the truth about what rationality requires. More generally, it seemed like agents

---

[39]There may have even been one of these loops in our original Self-Undermining argument, if you think that the move from p and "my situation contains testimony that p" to "I am rationally forbidden to believe p" requires a Closure-type step.

[40]Put in our earlier functional terms, they describe general features of the values of $\mathcal{R}$.

could receive evidence that permitted them to have rational, false beliefs about the requirements of rationality.

But self-undermining cases are importantly different from typical cases, and they show that the generalization from typical cases fails. Rational rules need to be properly restricted so as not to undermine themselves or each other. The result of those restrictions is that testimony contradicting the rational rules does not make it rationally permissible to believe falsehoods about the rules. Generally, an agent's total evidence will never all-things-considered support an *a priori* falsehood about the rational rules, because the rational rules are constructed such that no situation permits or requires a belief that contradicts them. There may be pieces of evidence that provide *some* reason to believe a falsehood about the rational rules, or evidence may be able to provide *pro tanto* support for such false beliefs. But the properly restricted rules will never make such false beliefs all-things-considered rational.

Now it may seem that what I've called the "proper" restrictions on rational rules are an overreaction. For example, we could adopt the following narrower restriction on Testimony:

> **Current-Situation Testimony:** If an agent's situation includes testimony that x, the agent is rationally permitted and required to believe that x—unless x contradicts an *a priori* truth about what rationality requires *in the agent's current situation*.

Current-Situation Testimony is more narrowly restricted than Properly Restricted Testimony because it bans testimony from requiring belief only when that testimony contradicts what the agent's *current situation* requires. Yet current-situation restrictions are still strong enough to prevent *akrasia* in the loop case. (Because p contradicts a fact about requirements in the agent's current situation, Current-Situation Closure would not require the agent to believe p.) Current-Situation Testimony is also of interest because it would be the rule endorsed by someone who accepted the Special Case Thesis but refused to accept its first generalization—the generalization that goes beyond mistakes about what's required in one's current situation to mistakes about what's required in other situations.

So what's wrong with Current-Situation Testimony, and why do we need the further step to Properly Restricted Testimony? Put bluntly, the problem with Current-Situation Testimony is that it's *ad hoc*. But we can fill out this charge more precisely, by connecting the rational rules we've been considering to such familiar epistemological notions as justification, evidence, and reasons.

I keep saying that the evaluations involved in our rational rules are all-things-considered evaluations. So in the original Self-Undermining case, if Anti-Akrasia (or Testimonial Parity) is true then the agent cannot be all-things-considered permitted to believe p. Plausibly, this means that the agent can't be all-things-considered *justified* in believing p, and that her evidence cannot all-things-considered *support* p. But that doesn't mean her evidence provides *no* support for p. And if we're going to grant that the testimony in Self-Undermining provides at least some *pro tanto* or *prima facie* justification for p, we need to tell a story about what outweighs or defeats that justification, creating an all-things-considered verdict consistent with Anti-Akrasia.

A similar story will be required for the loop cases. Even if we respond to those cases by adopting just Current-Situation Testimony, we're going to need to explain what offsets the justification testimony provides for false claims concerning what's

required in one's current situation. Similarly, if we accept the Special Case Thesis we'll need to explain what justificatory arrangment makes it rationally impermissible to believe an attitude is forbidden when it's actually required in one's current situation. Certainly if attitude A is required in an agent's situation, the agent will have support for that attitude. But that's different from having support for the proposition that A is required, or counter-support for the proposition that A is forbidden.

Ultimately, we need a story that squares Anti-Akrasia with standard principles about theoretical support and justification.[41] How is the justificatory map arranged such that one is never all-things considered justified in both an attitude A and the belief that A is rationally forbidden in one's current situation? The most obvious answer is that every agent possesses *a priori*, propositional justification for true beliefs about the requirements of rationality in her current situation.[42] An agent can reflect on her situation and come to recognize facts about what that situation rationally requires. Not only does this reflection provide her with justification to believe those facts; that justification is ultimately empirically indefeasible.[43]

I said this is the most obvious way of telling the kind of story we need; it's not the only way. But every plausible story I've been able to come up with is *generalizable*: it applies just as well to an agent's conclusions about what's rationally required in other situations as it does to conclusions about what's required in her current situation. For example, take the universal-propositional-justification story I've just described. However it is that one reflects on a situation to determine what it rationally requires, that process is available whether the situation is one's current situation or not. The fact that a particular situation is currently yours doesn't yield irreplicable insight into its *a priori* rational relations to various potential attitudes. So agents will not only have *a priori* propositional justification for truths about the rational requirements in their own situations; they will have *a priori* justification for true conclusions about what's required in any situation.

The generalizability of such stories makes it clear why the restriction in Current-Situation Testimony is *ad hoc*. Whatever keeps testimony from all-things-considered permitting false beliefs about one's *own* situation will also keep testimony from permitting false beliefs about *other* situations. This moves us from Current-Situation Testimony's narrow restriction to Properly Restricted Testimony's general restriction on false rational requirement beliefs. Properly Restricted Testimony then gives us our first generalization of the Special Case Thesis. Properly Restricted Testimony keeps testimony from providing rational permission to believe anything that contradicts an *a priori* rational-requirement truth—whether that truth concerns one's current situation or not. Parallel proper restrictions on other rational rules will prevent any rational permission to believe an attitude is forbidden when in

<hr>

[42]For discussion of positions similar to this one and further references, see (Field 2005) and (Ichikawa and Jarvis 2012, Chapter 7).

[43]Let me be clear what I mean by "ultimately indefeasible" here, because "indefeasible" is used in many ways. The story I'm imagining might allow that an agent's *a priori* propositional justifications for truths about rational requirements could be opposed by empirical evidence pointing in the other direction, empirical evidence that has some weight. But those propositional justifications are ultimately indefeasible in the sense that the empirical considerations will never outweigh them and make it all-things-considered rational for the agent to form false beliefs about what her situation requires.

fact that attitude is required. This holds whether or not the situation under consideration is one's own. And that's the first generalization of the Special Case Thesis.

We still need one more generalization for the full Fixed Point Thesis.[44] We also need to see why the bottom-up reading of Jane's case from Section 1 is preferable to the top-down view; both seem equally compatible with everything we've said so far. Curiously, both needs can be met by considering how the Special Case Thesis (with its first generalization) settles the peer disagreement debate.

<div align="center">

3

</div>

Suppose Greg and Ben are epistemic peers, in the sense that they're equally good at drawing rational conclusions from their evidence. Moreover, suppose that as part of their background evidence Greg and Ben know that they're peers in this sense. Now suppose that at t_0 Greg and Ben have the same total evidence E relevant to some proposition h, but neither has considered h and so neither has adopted a doxastic attitude towards it. For simplicity's sake I'm going to conduct this discussion in evidentialist terms (the arguments would go equally well on other views), so Greg and Ben's situation with respect to h just is their total relevant evidence E. Further suppose that for any agent with total relevant evidence E who adopts an attitude towards h, rationality requires that attitude to be belief in h. Now suppose that at t_1 Greg realizes that E requires believing h and so believes h on that basis, while Ben mistakenly concludes that E requires believing $\sim h$ and so (starting at t_1) believes $\sim h$ on that basis. (To help remember who's who: Greg does a $g$ood job rationally speaking, while Ben does $b$adly.)

At t_1 Greg and Ben have adopted their own attitudes towards h but each is ignorant of the attitude the other has adopted. At t_2 Greg and Ben discover their disagreement about h. They then have identical total evidence E', which consists of E conjoined with the facts that Greg believes h on the basis of E and Ben believes $\sim h$ on the basis of E. The question is what attitude Greg should adopt towards h at t_2.

A burgeoning literature in epistemology[45] examines this question of how peers should respond to disagreements in belief. Meanwhile peer disagreement about what to *do* (or about what intentions are required in a particular situation) has become a serious issue in moral theory, in part because of the troubling amount of such disagreement we seem to find in the actual world.[46] To keep a fixed case before our minds I'll focus on epistemological examples concerning what to believe in response to a particular batch of evidence, but my arguments will apply equally to disagreements about the intentions rationally required by a situation. To make the case even more concrete, I will sometimes suppose that in our Greg-and-Ben

---

[44]Why doesn't Properly Restricted Testimony—with its general refusal to permit false *a priori* beliefs about what rationality requires—give us the second generalization as well? If you look back at the Self-Undermining and loop cases we used to establish Properly Restricted Testimony, they both involved an agent's believing some attitude was forbidden when in fact it was required. So technically we haven't demonstrated any problems for cases in which the agent makes a weaker mistake about what rationality requires.

[45]Besides the specific sources I'll mention in what follows, (Feldman and Warfield 2010) is a collection of essays exclusively about peer disagreement. Other such collections are forthcoming.

[46]See, for example, (Enoch 2011, Ch. 8), (McGrath 2008), (Crisp 2007), (Sher 2007), (Wedgwood 2007, Sect. 11.3), and (Shafer-Landau 2003).

example E entails h. We might imagine that Greg and Ben are each solving an arithmetic problem, E includes both the details of the problem and the needed rules of arithmetic, and Ben makes a calculation error while Greg does not.[47] While disagreement cases are certainly not confined to such entailment cases, as far as I know every player in the debate is willing to accept entailment cases as a fair test of his view.

I will focus primarily on two responses to peer disagreement cases. The Split the Difference view (hereafter "SD") says that, in recognition of the fact that an epistemic peer has drawn the opposite conclusion about h, Greg is rationally required to suspend judgment about h.[48] The Right Reasons view (hereafter "RR") says that since Greg drew the rationally-required conclusion about h before discovering the disagreement, abandoning his belief in h at t_2 would be a rational mistake.

One might think that with the Fixed Point Thesis in place, RR is obviously correct. Here's a quick argument from Fixed Point to RR: At t_1 Greg comes to believe h on the basis of E, and also forms the belief that believing h is rationally required by E. (If we're imagining a calculation case, Greg thinks h is the right answer to the calculation and other answers are incorrect.) Since E does rationally require belief in h, the latter belief is true, and so rationally required by the Fixed Point Thesis. If at t_2 Greg follows SD, he will be making a rational mistake—in fact, two rational mistakes. First, he will be dropping the belief (h) that is rationally required by his evidence. Second, Greg will be abandoning a correct belief about what rationality requires. For surely what happens to Greg at t_2 (under SD) is that encountering Ben makes him doubt whether E did in fact support h (whether h reported the correct result of the calculation). Under SD Greg abandons the correct belief about what E requires, which is a rational mistake by lights of the Fixed Point Thesis.

But this argument is too quick. On the first putative mistake—dropping the belief in h that was required by his evidence—we need to be clear that Greg's evidence at t_2 (call it E') is different from the evidence E he had at t_1. E' includes all of E, plus additional information about Ben's conclusions. While we've stipulated that E requires belief in h, this doesn't show that E' does. The main dispute between SD and RR is whether E' still supports h, and hence whether Greg is still required to believe h at t_2. On Greg's second putative mistake—dropping at t_2 the true belief he had at t_1 about what E requires—abandoning a true belief about rationality is not necessarily condemned by the Fixed Point Thesis. There are two ways to suspend judgment about a proposition. Greg might at t_2 form the belief that there is a unique rationally required attitude towards h required by E, and that attitude is suspension of judgment. But I don't think that's what the SD defender imagines. I think that on SD, once Greg learns that his epistemic peer has drawn a different conclusion from him he doesn't know what to think about what E requires. Greg doesn't suspend judgment about h at t_2 because he thinks that's what E requires. Instead, he stops having a firm opinion about what E requires and suspends judgment about h because he doesn't know what E recommends.

---

[47]This is essentially the restaurant-bill tipping example from (Christensen 2007).

[48]SD is distinct from the "Equal Weight View" defended by Elga in (Elga 2007) and (Elga 2010). But for cases with particular features (including the case we are considering), Equal Weight entails SD. Since SD can be adopted without adopting Equal Weight more generally, I will use it as my target here.

The Fixed Point Thesis—and all the arguments we provided for it—suggest that it's a rational mistake to form a fixed, false opinion about what rationality requires in a given situation.[49] But that doesn't show that it's a rational mistake to refuse to form any opinion about what rationality requires in a particular situation out of a sense of total bewilderment.[50]

So we need a better argument that Greg should stick to his belief in h. Ironically, a good argument *for* RR can be developed from what I think is the best argument *against* RR. The anti-RR argument runs like this: Suppose for *reductio* that RR is correct and Greg shouldn't change his attitude towards h in light of the information that his peer reached a different conclusion on the same evidence. Now what if Ben was an epistemic superior to Greg, someone who Greg knew was much better at accurately completing arithmetic calculations? Surely Greg's opinion about h should budge somewhat once he learns that an epistemic superior has judged the evidence differently. Or how about 100 superiors? Or 1000? At some point when Greg realizes that his opinion is in the minority amongst a vast group of people who are very good at judging such things, rationality must require him to at least suspend judgment about h. But surely these cases are all on a continuum, so in the face of just one rival view—even a view from someone who's just an epistemic peer— Greg should change his attitude towards h somewhat, *contra* the recommendation of RR.[51]

Call this the Crowdsourcing argument against RR. It's a bit tricky to make out when we're working in a framework whose only available doxastic attitudes are belief, disbelief, and suspension of judgment—that framework leaves us fewer gradations to make the continuum case that if Greg should go to suspension in the face of some number of disagreeing experts then he should make at least *some* change in response to disagreement from Ben. But no matter, for all I need to make my case is that there's some number of epistemic superiors to Greg whose disagreement with him would make it rationally obligatory for him to suspend judgment about h. Because if you believe that, you must believe that there is some further, perhaps much larger number of epistemic superiors to Greg whose disagreement with him would make it rationally obligatory for him to believe $\sim h$. If you like, imagine the change happens in two steps, and with nice round numbers. First Greg believes h on the basis of E, and believes that E rationally requires him to do so. He then meets 100 experts who believe $\sim h$ on the basis of E, and believe that $\sim h$ is rationally required on evidence E. At this point Greg suspends judgment about h. Then he meets another 900 experts with the same opinion, and finally caves. Respecting their expertise, he comes to believe $\sim h$ and to believe that $\sim h$ is rationally obligatory on evidence E.

---

[49]The Fixed Point Thesis has to be about criticizing mistaken beliefs agents *do* form—as opposed to criticizing beliefs agents *fail* to form—to avoid Cognitive Capacity concerns.

[50]Note that these responses to the two putative mistakes are consistent. The SD defender may think that E' requires Greg at t_2 to suspend judgment both about h and about what attitude towards h is required by E.

[51]You might think of this as an argument *for* SD, but in fact it is merely an argument *against* RR. Kelly (2010, pp. 137ff) makes exactly this argument against RR, then goes on to endorse a Total Evidence View concerning peer disagreement that is distinct from SD. Whether this argument is an argument *for* anything in particular won't matter in what follows—though we should note that Kelly explicitly endorses the continuum point I just made.

If you have the intuition that got the Crowdsourcing argument going to begin with, that intuition should extend to the conclusion that faced with enough opposing experts Greg should come to believe both $\sim h$ and that this attitude is rationally obligatory on E. But if the Fixed Point Thesis is right, that proves too much: since (by supposition) E rationally requires belief in h, Greg is now making a rational error. So it is not rational for Greg to respect the experts in this way. And by the continuum idea, it's not rational for Greg to suspend judgment in the face of fewer experts to begin with, or even to budge in the face of disagreement from Ben his peer.

To get this result, we don't need the full Fixed Point Thesis; we just need the Special Case Thesis and the first generalization I suggested in section 2.2. Belief in h is required on E, and Greg now believes that such belief is forbidden. E isn't his current situation (since it's t_2, his current situation is E'), so we do need that first generalization. But that was the one we were able to establish in Section 2.3. Moreover, this argument for RR suggests a strategy for establishing the second generalization once we have the first. Suppose we have a view that permits an agent to make rational-requirement errors other than errors in which she takes something to be forbidden that's required. (These latter errors are the kind covered by Special Case with its first generalization.) Whatever kind of case motivates such permissions, we will be able to construct a more extreme version of that case in which the agent is indeed permitted to believe something's forbidden that's required. The strategy therefore leverages violations of the second generalization into violations of the Special Case Thesis with the first generalization, which we've already established. Many details remain about how exactly this strategy would go, but I believe it's a promising route to the full Fixed Point Thesis.

We now have an argument from the Fixed Point Thesis (or a slightly generalized Special Case Thesis) to the Right Reasons view about peer disagreement. It remains to consider objections to the RR view. But first I want to use the argumentative strategy we've just employed to dismiss the top-down reading of Jane's case from Section 1.

In Section 2.1 we dismissed the mismatch reading of Jane's case on the grounds that it conflicts with Anti-Akrasia. This left the top-down and bottom-up readings, neither of which obviously conflicts with the Fixed Point Thesis. Jane believes that her body of total evidence—which includes the proposition $p\&q$—rationally requires her to believe $\sim q$.[52] That might seem crazy if that total evidence were just $p \& q$, but the top-down theorist is imagining a version of the case in which Jane has other evidence in addition to $p \& q$, perhaps testimony from some authoritative source. The top-down theorist imagines that Jane's *total* evidence requires belief in $\sim q$, which means that Jane's belief about what's rational in her situation is true. So she is not obviously in violation of the Fixed Point Thesis.

The trouble comes when we imagine how that authoritative testimony might do its work. Suppose that Jane gains her evidence over time, and that at first her only evidence relevant to q was $p \& q$. Let's further suppose that if your total relevant evidence is just $p \& q$, you are indeed required to believe q. Now we imagine (along with the top-down theorist) that Jane gets some testimony from an authoritative source that, when combined with $p \& q$, gives her a bundle of total evidence that

---

[52]Here I'm continuing my simplifying evidentialist assumption that as far as rational requirements on doxastic attitudes go, an agent's situation just is her total evidence.

rationally requires belief in $\sim q$. The question is what the content of that testimony might have been. It seems clear that that testimony must somehow be about—or at least have implications for—what doxastic attitudes are rationally permissible on the evidence $p \,\&\, q$. If testimony has the kind of power to require Jane to believe $\sim q$ on her total evidence at the later time, it must also have the power to rationally require her to believe that belief in $\sim q$ is required (and belief in q is forbidden) on her initial evidence of just $p \,\&\, q$. But that *is* a falsehood about rational requirements—even by the top-down theorist's lights—and so represents a violation of the Fixed Point Thesis. (Or even just the Special Case Thesis with its first generalization.)

If the Fixed Point Thesis is true, misleading evidence at higher levels about what attitudes are required by evidence at lower levels does not "trickle down" to permit attitudes that otherwise would have been forbidden. SD and the top-down reading of Jane's case both fail because they are trickle-down theories. RR and the bottom-up reading are correct: if one's initial situation requires a particular attitude, that attitude is still required no matter how much misleading evidence one subsequently receives about what attitudes were permitted in the initial situation.

I now want to consider objections to RR. A number of precise, philosophical objections have been offered to the view; I think most of those have been adequately responded to elsewhere.[53] Yet there's an intuitive objection that immediately occurs to anyone when they first hear RR. RR says that if Greg drew the required conclusion from his evidence in the first place, he would be mistaken to abandon that conclusion when confronted with Ben's view. But how does Greg *figure out* that he's the one who should stick? Surely Ben thinks he's formed the right conclusion about h on E—so won't RR encourage him to (mistakenly) stick to his belief that $\sim h$?

The objection is a bit nebulous, and can be read many ways. In responding to any of those readings, it helps to consider an analogy. Suppose I defend the norm, "If you ought to ϕ, then you ought to perform any available ψ necessary for ϕ-ing." There may be many good objections to this norm, but here's a bad objection: "If I'm trying to figure out whether to ψ, how can I tell whether I ought to ϕ?" The norm in question is a conditional—it only applies to people meeting a certain condition. It is not the job of this norm to tell you (or help you figure out) whether you meet that condition. Similarly, it's no objection to this norm to say that if someone mistakenly thinks he ought to ϕ (when really he shouldn't), then his attempts to follow this norm may lead him to perform a ψ that he really shouldn't either. The norm says how agents should behave when they *actually* ought to ϕ, not when they *think* they ought to.

Similarly, RR is a conditional, describing what an agent is rationally required to do upon encountering disagreement *if* he drew the conclusion required by his evidence at an earlier time. It isn't RR's job to describe what Greg's initial evidence E requires him to believe; we have other rational rules (of entailment, of evidence, of perception, etc.) to do that. It also is no objection to RR that if Ben mistakenly

---

[53]I'm thinking especially of Elga's (2007) "bootstrapping" objection, which Elga thinks rules out any view other than SD. Kelly (2010, pp. 160ff.) shows that this objection applies only to a position on which both Greg and Ben should stick to their original attitudes (or something close to their original attitudes) once the disagreement is revealed. Thus bootstrapping is not an objection to RR or to Kelly's own Total Evidence View. (Though my "proves too much" objection to SD works against Total Evidence as well.)

thinks he meets its antecedent, his attempts to follow RR may lead him to adopt the wrong attitude towards h at t_2. In describing the case we stipulated that Ben was rationally required to believe h on the basis of E at t_1; Ben made a *rational error* when he concluded $\sim h$ instead. Any mistakes Ben then makes at t_2 from misapplications of RR are parasitic on his original miscalculation of what E rationally required at t_1. It shouldn't surprise us that an agent who misunderstands what's rationally required of him may go on to make further rational mistakes.

Perhaps the objection to RR involves some kind of Cognitive Reach concern: it's unreasonable to require Greg to stick to his beliefs at t_2 when it may not be obvious or accessible to him that he was the one who got things right. My response here is the same as it was to Cognitive Reach concerns about the Special Case Thesis.[54] The objection is motivated by the thought that in order for an attitude to be rationally required of an agent, the relevant relation between that attitude and the agent's situation must be sufficiently obvious or accessible. We stipulated in our example that at t_1 Greg and Ben are rationally required to believe h on the basis of E. In order for that to be true, the relevant relation between h and E (in the imagined case, an entailment) must have been sufficiently obvious or accessible to both parties at t_1. That obviousness or accessibility doesn't disappear when Greg gains more evidence at t_2; adding facts about what Ben believes doesn't keep Greg from recognizing h's entailment by E. So the facts needed for Greg to determine what RR requires of him are sufficiently obvious and accessible to him at t_2.

One might think that the extra information about Ben's beliefs contained in E' *defeats* what Greg knew at t_1—the extra evidence somehow destroys the all-things-considered justification Greg had for believing h at t_1. But that's just what's at issue between the RR-theorist and the SD-theorist: the former thinks E' still rationally requires Greg to believe E, while the latter does not. That E' contains defeaters for E's justification of h cannot be *assumed* in arguments between the two positions.

4

This essay began with logical omniscience. Examining formal epistemologists' struggles to remove logical omniscience requirements from their theories, we uncovered a duality relationship: any rational requirement—whether it be a requirement on beliefs or intentions, whether it be a requirement of attitudinal consistency or a constraint on inference—comes with particular propositions towards which agents are required (or forbidden) to adopt particular attitudes. Some of those propositions are propositions about rationality itself; the Fixed Point Thesis reveals that whenever there is a rational requirement, rationality also requires agents not to get the facts about that requirement wrong. This thesis concerns actual attitudes held by actual agents, not just agents who have been idealized somehow; it remains true whatever constraints we put in place on how many attitudes an agent can assign or how obvious a relation must be to generate rational requirements.

I established the Fixed Point Thesis through two arguments (No Way Out and Self-Undermining), each of which can be run using only Anti-Akrasia as a premise. I then showed that the Fixed Point Thesis has surprising consequences for agents'

---

[54]Once again (see note 33 above), I think the intuitive worry under consideration is available to both internalists and externalists in epistemology. Internalists are more likely to put the objection in terms of accessibility, while externalists are more likely to complain of insufficient obviousness.

responses to information about what's rational. If an agent has correctly determined what attitudes her situation requires, rationality forbids changing those attitudes in the face of evidence that she's made that determination incorrectly. Applied to peer disagreement cases, this implies the Right Reasons view on which an agent who's adopted the attitude required by her evidence is required to maintain that attitude even after learning that others have responded differently.

To my mind the strongest objection to the Fixed Point Thesis is not to offer some recondite philosophical argument but simply to deny its implications for disagreement on intuitive grounds. It seems preposterous that in the Crowdsourcing case Greg is required to stick to the (admittedly correct) conclusion of his calculations in the face of 1000 acknowledged mathematical experts telling him he's wrong.[55] If one wants to hold fast to that intuition, how should one respond to this essay's arguments?

One option is to deny Anti-Akrasia, the premise that got our arguments for the Fixed Point Thesis going.[56] Yet in addition to the general philosophical, intuitive, and historical unattractiveness of denying Anti-Akrasia, doing so in order to rescue Split the Difference (or a compromise like Kelly's Total Evidence View) raises two special problems—one motivational and one dialectical.[57] The first problem is that our intuition that Greg should concede at least some ground to the experts is motivated by a feeling that those experts offer him strong evidence that his initial conclusions were wrong, and that rationality requires Greg to respect that evidence. But if we are strongly wedded to the thought that rationality requires agents to respect their evidence, then it's bizarre to deny Anti-Akrasia. (There's a reason Feldman (2005) called his version of the Anti-Akrasia principle "Respect Your Evidence.") If we are committed to the rational importance of following one's evidence, shouldn't we require rational agents to be so committed as well? Anti-Akrasia forbids an agent from adopting an attitude that she takes her evidence to rule out.[58] If in our zealous defense of evidence from experts we deny Anti-Akrasia, we will allow agents to have no respect for their evidence at all.[59]

The second, dialectical problem is that dropping Anti-Akrasia can make it harder for the SD defender to defend her own view. One possible response to the Crowdsourcing case is to say that while Greg is indeed required to agree with the experts in his higher-order views—that is, he is reqiured to believe along with them that E requires belief in $\sim h$—he is nevertheless also required to maintain his original, first-order belief in h. The usual reply to this suggestion is that such a response would put Greg in a rationally unacceptable akratic overall state. But this reply is unavailable if one has dropped Anti-Akrasia. (In fact, Weatherson (ms) attacks Anti-Akrasia precisely so he can endorse such a mismatch view.) On the other

---

[55]A similar intuitive point against the Fixed Point Thesis can be made using Elga's (msa) hypoxia case. (See also (Christensen 2010).) Everything I say about Crowdsourcing in what follows applies equally well to hypoxia and other similar examples.

[56]One will also have to deny Testimonial Parity—an alternate premise for the Self-Undermining Argument—but I take it that someone who's gone so far as to deny Anti-Akrasia will be willing to deny that premise as well.

[57]Thanks to Stew Cohen and Russ Shafer-Landau for discussion on this approach.

[58]Here I am still working with the simplifying evidentialist assumption that an agent's situation just is her total evidence.

[59]Christensen (ms) has a particularly good discussion of SD theorists' motivational reliance on principles like Anti-Akrasia.

hand, we have seen that embracing Anti-Akrasia leads to the Fixed Point Thesis, which in turn requires Greg not to compromise with the experts.

So perhaps we should use Anti-Akrasia and the Fixed Point Thesis to explain away our Crowdsourcing intuitions. As I suggested earlier, experience from typical situations gives us good reason to think that something like Testimony is a requirement of rational belief. Properly Restricted Testimony matches that experience, but also accommodates baroque situations involving self-undermining. It turns out that—initial appearances notwithstanding—misleading testimony can't leave us with all-things-considered total evidence supporting false beliefs about the requirements of rationality. Perhaps our intuitions about testimony as evidence don't generalize in the ways we might originally have thought, and the experts' testimony in the Crowdsourcing case doesn't make Greg all-things-considered required to draw a false conclusion about what E requires.

But if we just can't shake the Crowdsourcing intuitions, we might hold that Crowdsourcing (and peer disagreement in general) presents a rational dilemma for Greg.[60] Notice that this would be consistent with the Fixed Point Thesis. In its most precise form, the Fixed Point Thesis says that no situation rationally permits an overall state containing *a priori* false beliefs about what situations rationally require. It is consistent with the thesis that for some situations *no* overall state is rationally permissible—in some situations no rationally-flawless state is available. One might think that in the Crowdsourcing case, the experts' evidence makes rationally flawed any overall state of Greg's that doesn't concede anything to the experts, while the Fixed Point Thesis draws on Anti-Akrasia considerations to make rationally flawed any overall state that does concede something to the experts. The result is that no rationally flawless overall state is available to Greg in the face of the experts' testimony, and we have a rational dilemma.

Again, all of this is consistent with the Fixed Point Thesis, so to insist that Greg is in a rational dilemma would not undermine any of the conclusion I have drawn in this essay.[61] We would still have my central claim that mistakes about rationality are mistakes of rationality; we would simply be admitting that those mistakes can sometimes be avoided only by offending rationality in other ways. But one might wonder what the *point* is in establishing the Fixed Point Thesis if all it tells us about peer disagreements is that the agents involved confront a rational dilemma.

Well, the Fixed Point Thesis and the arguments for it have other implications besides those for peer disagreement. Among other things, they make logical omniscience requirements more plausible and show that substantive conclusions can be drawn from purely structural premises (as noted in Section 2.2). But the point is well-taken. To respond to it we should consider more broadly why it's worthwhile to make evaluations of rational flawlessness that are not immediately tied to prescriptions. Here I should emphasize again that the evaluations we've been considering are evaluations of *real* agents' overall states, not the states of some mythical ideal

---

[60]This is the option Christensen (ms) defends.

[61]I have also tried to define Right Reasons very carefully so that it indicates a rational *mistake* if Greg *abandons* his belief in h at t_2—making RR consistent with the possibility that Greg is in a rational dilemma at t_2. If we define Split the Difference equally carefully, it may turn out that if peer disagreement poses a rational dilemma then both RR and SD are true! Yet I don't think this is the reading of SD that most of its defenders want. They tend to write as if splitting the difference with Ben squares Greg entirely with rationality's demands, leaving him in a perfectly permissible, rationally-flawless state.

agents. What value can it have to learn that such a state is rationally flawed? Consider Jane again, who derives $\sim q$ from $p \,\&\, q$ because she thinks that kind of inference is rationally required. Having rejected the top-down view, we can now affirm that Jane's belief state is rationally flawed. While this affirmation doesn't automatically dictate what Jane should believe going forward, it certainly affects prescriptions for Jane's beliefs. If the top-down theorists were right and there were no rational flaws in Jane's overall state, there would be no pressure for her to revise her beliefs and so no possibility of a prescription that she make any change.

When it comes to rational dilemmas, it can be very important to our prescriptive analysis to realize that a particular situation leaves no rationally-flawless options—even if that doesn't immediate tell us what an agent should *do* in the situation. A number of epistemologists[62] have recently analyzed cases in which an agent is misled about or unsure of what rationality requires in her situation (without having interacted with any peers or experts). Some have even proposed amendments to previously-accepted rational principles on the grounds that those principles misfire when an agent is uncertain what's required.[63] Meanwhile practical philosophers[64] have considered what happens when an agent is uncertain which intentions are required by her situation. Many of these discussions begin by setting up a situation in which it's purportedly rational for an agent to be uncertain—or even make a mistake—about what rationality requires. As in peer disagreement discussions, the authors eliminate various responses the agent might have to her situation by pointing out that those responses violate putative rational rules (logical consistency of attitudes, probabilistic constraints on credences, versions of Anti-Akrasia, etc.).

But now suppose that the moment the agent makes a mistake about what rationality requires (or even—if logical omniscience requirements are correct—the moment she assigns less-than-certainty to particular *a priori* truths), the agent has already made a rational error. Then it is no longer decisive to point out that a particular path the agent might take while maintaining the mistake violates some rational rule, *because no rationally flawless options are available to an agent who persists in such an error.* If we view a particular situation as a rational dilemma, determining the right prescription for an agent in that situation shifts from a game of avoiding rational rule violations to one of making trade-offs between unavoidable violations. That's a very different sort of normative task,[65] and the first step to engaging the sorts of norms-of-the-second best involved in sorting out a rational dilemma is to realize that you're in one.

There is one final way to accommodate the Crowdsourcing intuition (if one still wishes to do so in the face of theoretical considerations adduced here). We might maintain the Fixed Point Thesis, maintain RR, and claim that while Greg is both required and permitted to retain his h belief even in the face of the experts, something else about his epistemic situation changes. Earlier I mentioned that there are two ways to suspend belief about a proposition: suspension that one believes is the rationally required response to one's evidence versus suspension because one has no idea what to make of one's evidence. (Combinations of these positions are available

---

[62]Such as (Chalmers ms, Ch. 2), (Elga msa), (Hasan ms), (Schechter ms), (Weatherson ms), (Christensen 2010), and farther back (Foley 1990) and (Fumerton 1990).

[63]See, for example, criticisms of Rational Reflection in (Christensen 2010) and (Elga msb), and criticisms of single-premise closure in (Schechter ms).

[64]Including (Sepielli 2009), (Wedgwood 2007, Sect. 1.4), and (Feldman ms).

[65]Compare Rawls' (1971) distinction between ideal and nonideal theory.

as well.) Similarly, one might think that while rationality requires Greg's doxastic attitude towards h to remain fixed in response to the experts, something else could change. For example, that attitude might lose some justification, or at least possess less justification than it would have had otherwise. (Consider how Greg's justification in h would have changed had all those experts agreed with him!)

Which brings us back to credences, and logical omniscience. For simplicity's sake I've conducted our peer disagreement discussion entirely in terms of binary attitudes. But credences might display the same sort of ambiguity I've just discussed for suspension of belief. For example, I might have a credence of 0.5 that a coin will come up heads because I have very strong evidence that the coin is fair, or I might have a credence of 0.5 because I have no idea whether it's fair.[66]

My point is that on both the full-belief and partial-belief perspectives, a report of an agent's doxastic attitude towards a proposition may not include all there is to say about the agent's stance towards that proposition. If the Fixed Point Thesis is right and there is a logical omniscience requirement (at least over some set of accessible or sufficiently obvious cases), then in our imagined situation Greg is rationally required to have a credence of 1 that E entails h, and is required to be at least as confident of h as he is of E regardless of what any peers or experts say. But if we want to accommodate the intution that the experts should have *some* affect on Greg, perhaps we can do so in the other dimensions of his doxastic outlook—those not fully captured in a straightforward report of Greg's attitude towards h.[67]

References

Adler, J. E. (2002). *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Arpaly, N. (2000). On acting rationally against one's best judgment. *Ethics 110*, 488–513.

Balcerak Jackson, M. and B. Balcerak Jackson (ms). Reasoning as a source of justification. Unpublished manuscript.

Bergmann, M. (2005). Defeaters and higher-level requirements. *The Philosophical Quarterly 55*, 419–436.

Bjerring, J. C. (ms). Impossible worlds and logical omniscience: An impossibility result. *Synthese*. Forthcoming.

Broome, J. (1999). Normative requirements. *Ratio 12*, 398–419.

Chalmers, D. J. (ms). *Constructing the World*. Oxford: Oxford University Press. Forthcoming.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: The MIT Press.

Christensen, D. (2007). Epistemology of disagreement: the good news. *Philosophical Review 116*, 187–217.

---

[66]Defenders of imprecise credences will suggest that in the latter case I assign a credence range of something like $(0, 1)$. But even with ranges in play we can generate similar problems. For instance, we can imagine cases in which two agents rationally assign the same credence range to a given proposition, but one agent has more evidence than the other that that range is correct.

Christensen, D. (2010). Rational reflection. *Philosophical Perspectives 24*, 121–140.

Christensen, D. (ms). Epistemic modesty defended. Unpublished manuscript.

Cresswell, M. J. (1975). Hyperintensional logic. *Studia Logica: An International Journal for Symbolic Logic 34*, 25–38.

Crisp, R. (2007). Intuitionism and disagreement. In M. Timmons, J. Greco, and A. R. Mele (Eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*, pp. 31–39. Oxford: Oxford University Press.

Eells, E. (1985). Problems of old evidence. *Pacific Philosophical Quarterly 66*, 283–302.

Elga, A. (2007). Reflection and disagreement. *Noûs 41*, 478–502.

Elga, A. (2010). How to disagree about how to disagree. In R. Feldman and T. A. Warfield (Eds.), *Disagreement*, pp. 175–186. Oxford University Press.

Elga, A. (msa). Lucky to be rational. Unpublished manuscript.

Elga, A. (msb). The puzzle of the unmarked clock and the new rational reflection principle. Unpublished manuscript.

Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Feldman, F. (ms). What to do when you don't know what to do. Unpublished manuscript.

Feldman, R. (2005). Respecting the evidence. *Philosophical Perspectives 19*, 95–119.

Feldman, R. and T. A. Warfield (Eds.) (2010). *Disagreement*. Oxford: Oxford University Press.

Field, H. (2005). Recent debates about the a priori. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, pp. 69–88. Oxford University Press.

Foley, R. (1990). Fumerton's puzzle. *Journal of Philosophical Research 15*.

Fumerton, R. (1990). *Reasons and Morality: A Defense of the Egocentric Perspective*. Ithaca, NY: Cornell University Press.

Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese 140*, 97–119.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (Ed.), *Testing Scientific Theories*, pp. 99–132. Minneapolis: University of Minnesota Press.

Gibbons, J. (2006). Access externalism. *Mind 115*, 19–39.

Hasan, A. (ms). A puzzle for analyses of rationality. Unpublished manuscript.

Ichikawa, J. and B. Jarvis (2012). *The Rules of Thought: A Rationalist Proposal*. Oxford University Press. (Forthcoming in fall 2012.).

Kant, I. (1974). *Logic*. New York: The Bobbs-Merrill Company. Translated by Robert S. Hartman and Wolfgang Schwarz.

Kelly, T. (2010). Peer disagreement and higher-order evidence. In R. Feldman and T. A. Warfield (Eds.), *Disagreement*, pp. 111–174. Oxford University Press.

McGrath, S. (2008). Moral disagreement and moral expertise. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, Volume 3, pp. 87–108. Oxford: Oxford University Press.

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scanlon, T. M. (2003). Metaphysics and morals. *Proceedings and Addresses of the American Philosophical Association 77*, 7–22.

Schechter, J. (ms). Rational self-doubt and the failure of closure. Forthcoming in *Philosophical Studies*.

Schroeder, M. (2008). Having reasons. *Philosophical Studies 139*, 57–71.

Sepielli, A. (2009). What to do when you don't know what to do. *Oxford Studies in Metaethics 4*, 5–28.

Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.

Sher, G. (2007). But i could be wrong. In R. Shafer-Landau (Ed.), *Ethical Theory: An Anthology*, pp. 94–102. Oxford: Blackwell Publishing Ltd.

Smithies, D. (2011). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research*.

Weatherson, B. (ms). Do judgments screen evidence? Unpublished manuscript.

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.

Williams, B. (1981). Internal and external reasons. In *Moral Luck*, pp. 101–113. Cambridge University Press.