9. Rationality's Fixed Point (or: In Defense of Right Reason)

Michael G. Titelbaum

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. (We'll see more reasons for this restriction below.) 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 farther than that. As I build my case for the thesis, we'll see how far we can make it extend.

[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.

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 to specify particular 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 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 total evidence about rational requirements.

Finally, the Fixed Point Thesis has implications for how one's higher-order beliefs (beliefs about what's rational in one's situation) should interact with one's first-order beliefs. Thus it has consequences for the peer disagreement debate in epistemology. Most philosophers think that in the face of disagreement with an equally rational, equally informed peer an agent should conciliate her opinions. Yet the Fixed Point Thesis implies that whichever peer originally evaluated the shared evidence correctly should stick to her guns.

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 apply the Fixed Point Thesis to higher-order reasoning and peer disagreement, and defend the thesis from arguments against it.

1. LOGICAL OMNISCIENCE

I first became interested in the Fixed Point Thesis while 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 agents' degrees of belief. One of those axioms (usually called Normality) assigns a value of 1 to every logical truth. In Bayesian epistemology this entails something like a rational requirement that agents assign certainty to all logical truths. Logical omniscience in some form is also a requirement of such formal epistemologies as ranking theory and AGM theory.

[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 word we 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' (1988) *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.

Logical omniscience requirements provoke four major objections:

- There are infinitely many logical truths. An agent can't adopt attitudes toward 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 not to recognize them as such and assign the required certainty. (Call this the Cognitive Reach objection.)
- Rational requirements are requirements of consistency *among* attitudes toward propositions. They do not dictate particular attitudes toward *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 toward rational requirements instead of attitudes toward logical truths.) The first two objections are rather straightforwardly met. For Cognitive Capacity, one need only interpret the relevant logical omniscience requirements as taking the form "*If* one takes an attitude toward a logical truth, *then* one should assign certainty to it." Logical omniscience then does not require that attitudes be taken toward 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 toward particular propositions (that is, if attitudes toward 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

[4] Alan Hájek first brought 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. (See e.g. Hegel (1975: pp. 75ff.).)

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 the formal epistemologies I've mentioned, 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]

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]

So special requirements on attitudes toward logical truths are not the sole province of *formal* epistemologies. Still, we can learn about such requirements by observing what happens to formal theories when the requirements are lifted. Formal theories don't require logical omniscience because formal theorists like the requirement; logical omniscience is a side-effect of systems

[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 on Gaifman's approach.

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

capturing the rational requirements theorists are 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 humans, this model allows an agent to be less than certain that if all humans are mortal then the human Socrates is as well. But in that model it may also be the case that *Ms* no longer confirms $(\forall x)Mx$, one of the basic confirmation relations we build Bayesian systems 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 human Socrates is mortal than she is that all humans 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 place constraints on consistency and inference, and so would be vitiated entirely.

What does logical omniscience have to do with this essay's main topic—attitudes toward truths about rational requirements? In general, a rational requirement on consistency or inference often stands or falls with a requirement on attitudes toward a particular proposition.[10] I call such a proposition a "dual" of the requirement on consistency or inference.[11] Logical omniscience

[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 $P(S|A) > P(S|{\sim}A)$, 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 possible 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 (2013).)

[10] Balcerak Jackson and Balcerak Jackson (2013) offer another nice example of this phenomenon. In classical logic an agent who can rationally infer y from x can also 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 $x \supset y$. (See also Brandom's (1994: Ch. 2) position that material conditionals just *say* that particular material inferences are permitted.)

[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.

requirements reveal logical truths to be duals of rational requirements—if an agent is not required to take a special attitude toward a particular logical truth, other potential requirements on her reasoning fall away as well. The Fixed Point Thesis affirms that each rational requirement also has a dual in the proposition expressing that requirement.[12] If rationality permits an agent to disbelieve an a priori proposition describing a putative rational requirement, the putative requirement is not a genuine one.[13]

Of course I need to *argue* for this thesis, and I will begin to do so soon. But first I should clarify my commitments coming out of this phase of the discussion. In what follows I will be agnostic about whether Cognitive Capacity and Cognitive Reach are good objections to theories of rationality. The arguments and theses advanced will be capable of accommodating these concerns, but will not be committed to their having probative force. The Fixed Point Thesis, for example, requires agents *not* to have *false* beliefs about rational requirements (instead of requiring agents *to* have *true* beliefs) so that no infinite belief set is required. Similarly, each argument to come will be *consistent* with limiting rational requirements on an agent's beliefs to what is sufficiently obvious or accessible to her. But those arguments will not *require* such limitations, either.

2. THE AKRATIC PRINCIPLE

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 simplicity's sake we'll stick to just belief, disbelief, and suspension of judgment) 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 example, there might be a (wide-scope) requirement of instrumental rationality that negatively evaluates any overall state

[12] I say "also," but on some understandings of logical truth the Fixed Point Thesis entails a logical omniscience requirement. Kant (1974) took logical truths to express 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 truths about rational requirements.

[13] I am *not* suggesting here 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 an agent has explicit higher-level views, and then argue that in such cases the requirements at different levels match.

that includes an intention to ϕ, a belief that ψ-ing is necessary for ϕ-ing, and an intention not to ψ.[14]

Rules of rationality require or permit certain kinds of overall states. But which states are permitted for a particular agent at a particular time may depend on various aspects of that agent's circumstances. Different philosophical views take different positions here. An evidentialist might hold that which doxastic attitudes figure in the overall states permitted an agent depends only on that agent's evidence. One natural development of this view would then be that the list of rationally permitted overall states for the agent (including both beliefs and intentions) varies only with the agent's evidence and her desires. But we might think instead that which intentions appear in permitted states depends on an agent's reasons, not on her desires. Or if we want to deny evidentialism, we might suggest that an agent's beliefs in the past influence which doxastic attitudes appear in the overall states permitted to her in the present.[15]

To remain neutral on these points I will assume only that whatever the true theory of rationality is, it may specify certain aspects of an agent's circumstances as relevant to determining which overall states are rationally permitted to her. 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 possessing that overall state in that situation. This is meant to be an *evaluative* excercise, which need not immediately lead to prescriptions—I am not suggesting a rational rule that agents ought only adopt rationally flawless states. In Section 7 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 to say sometimes that a particular overall state is the only flawless state available in a situation. But English offers no concise, elegant way to say things like that, especially when we want to put them in verb phrases and the like. So I will repurpose a terminology already to hand for describing states that satisfy all the principles of a kind and states that uniquely satisfy the principles of that kind: I will describe an overall state

[14] For the "wide-scope" terminology see Broome (1999). One might think that some requirements of practical rationality involve not just an agent's intentions but also her actions. In that case one would have to include actions the agent is in the process of performing at a given time in her overall state along with her attitudes. For simplicity's sake I'm going to focus just on rational evaluations involving beliefs and intentions.

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

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."[16]

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 attitude 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.

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 engaged in a purely evaluative project.[17] 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 worthwhile to consider what kinds of rational flaws may arise in a real agent's attitude set. Finally, my rational evaluations are all-things-considered 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.[18] 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 certain overall states are rationally required or permitted in particular

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

[17] When we get around to making *arguments* about what's required and permitted, it may appear that I'm assuming a substantive deontic logic (in particular something like Standard Deontic Logic) to make those arguments go through. But that will be a false appearance due to my idiosyncratic use of typically prescriptive language. Given the definitions I'm using for "required," "permitted," etc. all of my arguments will go through using only classical first-order logic (with no special deontic axioms or inference rules).

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

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** No situation rationally permits an a priori false belief about which overall states are rationally permitted in which situations.

I will argue to this thesis from a premise we can state as follows:

> **Akratic Principle** 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.

The Akratic Principle 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 to cases in which the agent has a belief *A* while believing that belief is forbidden in her situation.[19] The principle 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. So the Akratic Principle 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*.[20]

The irrationality of practical *akrasia* has been discussed for centuries (if not millennia), and I take it the overwhelming current consensus endorses the Akratic Principle 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 the principle's application to the belief case it is well worth reading Feldman's defense.[21] (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 the Akratic Principle for beliefs could after a quick logical step find herself

[19] I also take the Akratic Principle to apply to cases in which *A* is a combination of attitudes rather than a single particular attitude. While I will continue to talk about *A* as a single attitude for the sake of concreteness, any "*A*" in the arguments that follow can be read either as a single attitude or as a combination of attitudes.

[20] 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 the Akratic Principle indicates that if the agent adopts the rationally required attitude then his 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 (2000, p. 491) explicitly concedes the presence of that rational flaw.

[21] 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.

with a Moore-paradoxical belief of the form "x, but it's irrational for me to believe x."[22]

Still, objections to the Akratic Principle (in both its theoretical and practical applications) are available.[23] One important set of objections focuses on the fact that an agent might be mistaken about aspects of her current situation, or about aspects of her current overall state. Given my neutrality about the contents of situations, I cannot assume that all aspects of situations are luminous to the agents in those situations. So we might for instance have a case in which an agent believes that p, believes that it's rationally forbidden to believe something the negation of which one has believed in the past, actually did believe $\sim p$ in the past, but does not remember that fact now. I also do not want to assume that an agent is always aware of every element in her overall state.[24] So we might have a case in which an agent believes attitude A is rationally forbidden, possesses attitude A, but does not realize that she does. Or she might believe that attitudes meeting a particular description are forbidden, yet not realize of an attitude she has that it meets that description.

Rational evaluations in such cases are subtle and complex. The Akratic Principle might seem to indict the agent's overall state in all these cases, and I don't want to be committed to that. I have tried to formulate the principle carefully so as to apply only when an agent has the belief that her current situation, *described as her current situation*, rationally forbids a particular attitude. But that formulation may not handle all complications involving multiple descriptions of situations, and it certainly doesn't handle failures of state luminosity. Frankly, the best response to these objections is that while they are important, they are tangential to our main concerns here. For every case I will construct and every debate about such cases I will consider, that debate would remain even if we stipulated that the agent is aware of all the relevant situational features and of all her own attitudes (under whatever descriptions are required). So I will consider such stipulations to be in place going forward.[25]

[22] See Smithies (2012) for further discussion of such paradoxical statements.

[23] One objection we can immediately set aside is that of Audi (1990). Audi objects to the claim that if an agent judges better alternatives to an action to be available, then it's irrational for her to perform that action. But this is a narrow-scope claim, not our wide-scope Akratic Principle. We can see this in the fact that Audi's objection focuses exclusively on evaluating the *action*, and turns on a case in which the agent's judgment is mistaken. Audi does not investigate negative rational evaluations of that judgment, much less broader negative evaluations of the overall state containing both the judgment and the intention to perform the action, or of the agent herself. That his objection does not apply to such broader evaluations comes out when Audi writes, "I … grant that incontinence counts against the rationality of *the agent*: one is not fully rational at a time at which one acts incontinently" (1990: p. 80, emphasis in original). (For further analysis of Audi's objection see Brunero (2013: Section 1).)

[24] For the sorts of reasons familiar from Williamson (2000).

[25] Williamson (2011) uses an example involving an unmarked clock to argue that "It can be rational for one to believe a proposition even though it is almost certain on one's evidence that it is not rational for one to believe that proposition." While that is not quite a direct counter-example to the Akratic Principle, it can easily be worked up into one. (See, for instance, Horowitz (2013, Section 6).) However Williamson's example is explicitly set up to make it

Before moving on, however, I should note that these objections to the Akratic Principle bring out further reasons why we need the a priori rider in the Fixed Point Thesis. An agent might have a false belief about what's required in her situation because she mistakes the content of that situation. She might also falsely believe that her current state is rationally permitted in her current situation because she is incorrect about what her overall state contains. But neither of these false beliefs necessarily reveals a rational mistake on the agent's part. Each of them is really a mistake about an a posteriori fact—the contents of her situation or of her overall state.

So what kind of false belief *is* rationally forbidden by the Fixed Point Thesis? One way to see the answer is to think of rational requirements as describing a function $\mathcal{R}$. Reading an overall state as just a set of attitudes, we can think of $\mathcal{R}$ as taking each situation S to the set $\mathcal{R}(S)$ of overall states that would be rationally flawless for an agent to hold in S.[26] The Fixed Point Thesis would then hold that there do not exist a situation S', a situation S, and an overall state $O \in \mathcal{R}(S)$ such that O contains a false belief about the values of $\mathcal{R}(S')$. In other words, no situation permits an agent to have false beliefs about which overall states $\mathcal{R}$ permits in various situations. This formulation isolates out issues about whether an agent can tell that her current situation is S and her current overall state is O. Further, I take it that facts about the values of $\mathcal{R}$ are a priori facts.[27] So this formulation clarifies why a priori false beliefs figure in the Fixed Point Thesis.[28]

These complications aside, there's a much more intuitive objection to the Akratic Principle available. Weatherson (ms) presents this objection—and its underlying commitments—in a particularly clear fashion.[29] He begins with an example:

unclear to the agent what her evidence is. So I read Williamson's case as a failure of situational luminosity, and hence will set it aside.

[26] Notice that overall states can be "partial," in the sense that they don't contain a doxastic attitude toward every proposition or an intention concerning every possible action. This reflects my earlier response to Cognitive Capacity that rationality need not require agents to take attitudes toward everything.

[27] Even if situations 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 rationally permit which overall states. They will take the form "*if* the empirical facts are such-and-such, *then* rationality requires so-and-so."

[28] There are still some complications lurking about states and situations under disparate descriptions. For instance, we might think that the sentence "my current overall state is in $\mathcal{R}(S)$" (for some particular S) expresses a "fact about the values of $\mathcal{R}(S)$" that an agent could get wrong because she misunderstands her current state. I'm really picturing $\mathcal{R}$ taking as inputs situations described in some canonical absolute form (no indexicals, no *de re* locutions, etc.) and yielding as outputs sets of states described in a similar canonical form. The Fixed Point Thesis bans mistakes about which canonically described situations permit which canonically described states, without addressing mistakes about which non-canonically described situations/states are identical to the canonically described ones. While the details here are complex, I hope the rough idea is clear.

[29] Weatherson ultimately wants to deny a version of the Akratic Principle for the theoretical case. But he gets there by first arguing against a version for the practical case, and then drawing an analogy between the practical and the theoretical. (For another example similar to Weatherson's Kantians, see the Holmes/Watson case in Coates (2012).)

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 . . . .
>
> 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. (p. 12)

There are a number of differences between our Akratic Principle and the principle Weatherson is attacking. First, we are considering *intentions* while Weatherson considers 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 *shouldn't* have, while we're considering what combinations of attitudes would be *rationally flawed* for Frances to have. Can Weatherson's Kantians example be used to argue against our Akratic Principle, concerning rationally flawed overall states?

When we try to use Kantians 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. We might insist that if Frances has enough reason to believe that Kantian ethics is true, then Frances is rationally forbidden to lie to the axeman at the door. (And thus is not required to form an intention she believes is rationally forbidden.) Or, going in the

other direction, we 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. While these two responses move in opposite directions,[30] each denies that the case Weatherson has described (as transposed into rationality-talk) is possible.

These responses also bring out something odd about Weatherson's reading of the Kantians 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.[31] 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 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.

Weatherson cannot just stipulate in the Kantians 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 the Akratic Principle.[32]

There is another kind of case in which what an agent should do might diverge from what she should believe she should do. I suggested above that when testimony offers normative advice, any reason to believe that testimony can also be a reason to obey it, and vice versa. Yet we can have cases in

[30] In Section 5 we will give the positions that engender these responses names. In the terminology of that section, the first response would be popular with "top-down" theorists while the second belongs to a "bottom-up" view.

[31] I am not taking a position here on whether testimony is 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.

[32] 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 the Akratic Principle 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 7.

which certain reasons bear on behavior but not on belief. To see this possibility, consider Bernard Williams's famous example (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 were she *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 may provide an agent with reason to behave in particular ways if that fact is appropriately related to her desires.[33] 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 cognitively local information; what she has reason to *do* may be influenced by nonlocal facts.[34] And if we think that what 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.[35] 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. What one is *rationally* required to do or believe depends only

[33] Williams (1981: p. 103) writes, "[Agent] A may be ignorant of some fact such that if he did know it he would, in virtue of some element in [his subjective motivational set] S, be disposed to ϕ: we can say that he has a reason to ϕ, though he does not know it. For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that A would have a reason to ϕ if he knew the fact." Notice that whether the unknown fact counts as a reason for the agent depends on how relevant that fact is to the agent's actions given his motivations, *not* how cognitively local the fact is to the agent.

[34] 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 her reasons for higher-order beliefs would be identical. 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 the Akratic Principle for the theoretical case (in which A is a doxastic attitude) than for the practical case (in which A is an intention).

[35] 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.)

on what's cognitively local—that's what made Cognitive Reach a plausible objection. As long as the normative notion featured in the Akratic Principle is rational requirement, Williams-style cases don't generate counterexamples to the principle.

Once more this discussion of potential counterexamples to the Akratic Principle 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 the Akratic Principle, which concerns rational requirements. As we've just seen, that principle may be less plausible for other normative notions; for instance, Williams-style cases might undermine an Akratic Principle for reasons. But for any normative notion for which an analogue of the Akratic Principle holds, I believe I could run my arguments for a version of the Fixed Point Thesis featuring that normative notion. For normative notions for which a version of that principle is not plausible, I do not know if a Fixed Point analogue holds.

3. NO WAY OUT

I will now offer two arguments 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. Yet the Special Case Thesis is a good place to start, as many people inclined to deny the Fixed Point Thesis will be inclined to deny its application to this special case as well.[36]

While the Special Case Thesis may look a lot like the Akratic Principle, they are distinct. The Akratic Principle concerns the rational permissibility of an agent's assigning two attitudes at once. The Special Case Thesis concerns an agent's assigning a particular attitude when a particular rational requirement is in place. Yet despite this difference one can argue quickly from the principle to the thesis, and do so in multiple ways. I call my first argument from one

[36] For example, someone with Weatherson's inclinations might read Kantians as a case in which intending to lie is required, yet Frances is permitted to believe intending to lie is forbidden. If that reading were correct, Kantians would be a counterexample to the Special Case Thesis.

to the other No Way Out; it 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 the belief that A is rationally forbidden to her. 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 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 the belief that A is forbidden in the current situation. By the Akratic Principle this state is not rationally permissible, contrary to supposition once more. This completes our *reductio*. The Akratic Principle entails the Special Case Thesis.

It's surprising that the Special Case Thesis is so straightforwardly derivable from the Akratic Principle. Part of the surprise comes from deriving something controversial (if not downright counterintuitive) from something that the large majority of philosophers believe. 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.[37] 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, the Akratic Principle 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.[38]

[37] 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" (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 have to do with *rationality*, but by Scanlon (2003) he ceased to rely on that assumption.

[38] A similar move from structural to substantive occurred in my earlier argument from Confidence to the conclusion that logical truths require maximal credence. One might object that the No Way Out argument does not move solely from structural premises to a substantive conclusion, because that argument begins by assuming that there is at least one situation in which an attitude A is rationally required (which seems to involve a presupposed substantive constraint). I think that objection is harder to make out for the Confidence argument, but even with No Way Out a response is available. As I suggested in note 19, we can read "A" throughout the argument either as an individual attitude or as a combination of attitudes. Since structural constraints are requirements on combinations of attitudes, we can therefore run No Way Out for a case built strictly around structural assumptions. For a thorough presentation

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"), which is worrying because many epistemologists consider rationality an internalist notion.[39] But this 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 total evidence about rational requirements is impossible. The No Way Out argument moves from a premise about consistency requirements among an agent's attitudes (the Akratic Principle) to a strong conclusion about what can be substantively supported by an agent's evidence.

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 agent takes an attitude that's *required* to be *forbidden*. To reach the full Fixed Point Thesis, we would have to generalize the Special Case Thesis in two ways:

(1) to mistakes besides believing that something required is forbidden; and

of such cases and an explicit derivation of the substantive from the structural, see Titelbaum (2014).

[39] Of course, Cognitive Reach concerns need not be exclusive to (epistemological) internalists. 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.

(2) to mistakes about what's rationally required by situations other than the agent's current situation.

As an example of the first generalization, we would for example have to treat cases in which an attitude is rationally *forbidden* for an agent but the agent believes that attitude is *required*. This generalization is fairly easy to argue for, on the grounds that any well-motivated, general epistemological view that rationally permitted agents to have a belief at odds with the true requirements of rationality in this direction would permit agents to make mistakes in the other direction as well. (Any view that allowed one to believe something *forbidden* is *required* would also allow one to believe something *required* is *forbidden*.) Yet we already know from the Special Case Thesis that believing of a required attitude that it's forbidden is rationally impermissible. This rules out such epistemological views.[40]

The second generalization, however, is more difficult to establish. I'll argue for it by first presenting another route to the Special Case Thesis.

4. SELF-UNDERMINING

One strong source of resistance to the Fixed Point Thesis is the intuition that if an agent has the right kind of evidence—testimony, cultural indoctrination, 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 land 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

[40] In Section 6 I'll argue for another instance of the first generalization, one in which the mistake made about what's rational is less extreme than thinking what's required is forbidden (or vice versa).

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.[41]

The Self-Undermining Argument shows by *reductio* that Testimony cannot express a true general rule of rationality. Begin by supposing Testimony is true, then suppose that an agent receives testimony containing the following proposition (which I'll call "*t*"):

> 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 *t*. So suppose the agent is in that rationally permitted state. Since the agent believes *t*, she believes that it's rationally impermissible to believe testimony. She learned *t* from testimony, so she believes that belief in *t* is rationally forbidden in her situation. But now her overall state includes both a belief in *t* and a belief that believing *t* is rationally forbidden. By the Akratic Principle, the agent's state is rationally impermissible, and we have a contradiction. The Akratic Principle entails 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 *t*, she should stop viewing that source as an authority. (Anyone who says something as crazy as *t* 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, a source can't qualify as an authority by virtue of agents' being rationally required and permitted to believe what he 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.

[41] As stated, Testimony applies only to an agent's beliefs. Yet following on some of the points I made in response to Weatherson's Kantians argument in Section 2, we could create a general testimony norm to the effect that whenever testimony recommends a particular attitude (belief or intention), rationality permits and requires adopting that attitude. The arguments to follow would apply to this generalized norm as well.

Whatever those criteria are, we can stipulate that the source providing testimony that t in the Self-Undermining Argument has met those criteria. Then the claim that the agent should stop treating her source as an authority the moment that source says t 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 t.

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 exclusion of just such testimony. 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.

Restricted Testimony performs exactly like Testimony in the everyday cases that lend Testimony intuitive plausibility. But the added restriction inoculates the rule against Self-Undermining; it stops that argument at its very first step, in which the agent's receiving testimony that t makes it permissible for her to believe t. t 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.[42] Thus the restriction in Restricted Testimony keeps testimony that t from rationally permitting or requiring the agent to believe t.[43]

There's nothing special about Testimony as a rational rule here—we're going to want similar restrictions on other rational rules to prevent Self-Undermining. For example, we might have the following:

[42] If we read both t 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 t and Restricted Testimony contradict each other.

[43] One might think that the move from Testimony to Restricted Testimony is unnecessary, because a realistic version of the Testimony rule would exempt testimony from permitting belief when defeaters for that testimony are present. If t—or any other proposition that similarly contradicts the Testimony rule—counts as a defeater for any testimony that conveys it, then a Testimony rule with a no-defeaters clause will not be susceptible to Self-Undermining. Yet if one could successfully establish that such propositions always count as defeaters, then the no-defeaters Testimony rule would come to the same thing as the Restricted Testimony rule (or perhaps a Restricted Testimony rule with a no-defeaters clause of its own). And no-defeaters Testimony would still be susceptible to the looping problem I'm about to describe for Restricted rational rules.

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 place.

While these Restricted principles may seem odd or ad hoc, they have been seriously proposed, assessed, and defended in the epistemology literature—see Weiner (2007), Elga (2010), Weatherson (2013), and Christensen (2013).[44] But that literature hasn't noticed that restricting rules from *self*-undermining doesn't solve the problem. Rational rules must not only include 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 t is true or she has no hands (where t is as before). By Restricted Testimony, there is a state rationally permitted in that agent's situation in which she believes that either t is true or she has no hands. (Notice that this disjunctive 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 in t (which, while it contradicts Restricted Testimony, does not contradict Restricted Closure). But t indicates that the agent is rationally forbidden to believe that either t is true or she has no hands, and we can complete our argument as before by the Akratic Principle.

At no point in this argument does one of our restricted rational rules dictate that a belief is required or permitted that logically contradicts *that rule*. Instead we have constructed a loop in which no rule undermines itself but together

[44] This discussion takes off from a real-life problem encountered by Elga. Having advocated a conciliatory "Split the Difference" position on peer disagreement like the one we'll discuss in Section 6, 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 itself. More real-life self-undermining: The author Donald Westlake once joked that when faced with a t-shirt reading "Question Authority," he thought to himself "Who says?" And then there's 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."

the rules wind up undermining each other.[45] 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.[46] 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 intuitively thought that even when this happens, 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 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 structured 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 provide prima facie

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

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

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 restricted less than Properly Restricted Testimony because it prevents testimony from permitting belief only when that testimony misconstrues what the agent's *current situation* requires. Yet current-situation restrictions are still strong enough to prevent *akrasia* in the loop case. (Because t contradicts a fact about requirements in the agent's current situation, Current-Situation Closure would not require the agent to believe that t.) 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 second 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.[47]

With that said, I don't find Current-Situation Testimony at all plausible—it's an egregiously ad hoc response to the problems under discussion. Yet by investigating *in exactly what way* Current-Situation Testimony is ad hoc we can connect 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. If the Akratic Principle is true, the correct rational rules will be restricted in *some* way to keep an agent who receives testimony that t from being all-things-considered permitted to believe t. Plausibly, this means that the agent won't be all-things-considered justified in believing t, and that her total evidence won't all-things-considered support t. But that doesn't mean that *none* of her evidence will provide *any* support for t. And if we're going to grant that testimony can provide *pro tanto* or *prima facie* justification for believing t, we need to tell a story about what outweighs or defeats that justification, creating an all-things-considered verdict consistent with the Akratic Principle.

Similarly, if we respond to the loop cases by moving to Current-Situation Testimony (without going all the way to Properly Restricted Testimony), we still need to explain what offsets the incremental justification testimony provides for false claims concerning what's required in one's current situation. And if we accept the Special Case Thesis, we need to explain what justificatory arrangement makes it impermissible to believe that a rationally required

[47] I am grateful to Shyam Nair for discussion of Current-Situation rules.

attitude is forbidden. Certainly if attitude *A* is required in an agent's situation, the agent will have support for *A*. 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 the Akratic Principle with standard principles about belief support and justification. 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.[48] An agent can reflect on her situation and come to recognize facts about what that situation rationally requires. Not only can this reflection justify her in believing those facts; the resulting justification is also empirically indefeasible.[49]

I said this is the most obvious way to tell the kind of story we need; it is 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 situations other than her own 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 irreproducible 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.[50]

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 weak restriction to Properly Restricted Testimony's general restriction on false rational-requirement beliefs. Properly Restricted Testimony and the other Properly Restricted rules

[48] For discussion of positions similar to this one and further references, see Field (2005) and Ichikawa and Jarvis (2013: Chapter 7).

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

[50] Another available backstory holds that everything I've just said about empirically indefeasible propositional justification is true for all a priori truths—there's nothing special about a priori truths concerning rational requirements. Clearly *that* story is generalizable, but assessing it is far beyond the scope of this essay.

then give us our second 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 prevent any rational permission to believe an attitude is forbidden that is in fact is required. This holds whether or not the situation under consideration is one's own. And that's the second generalization of the Special Case Thesis.

5. THREE POSITIONS

My argument from the Akratic Principle to the (full) Fixed-Point Thesis is now complete. It remains to consider applications of the thesis and objections to it. To understand the thesis's consequences for higher-order reasoning, we'll begin with an example.

Suppose Jane tells us (for some particular propositions p and q) that she believes it's not the case that either the negation of p or the negation of q is true. Then suppose Jane tells us she also believes the negation of q. $\sim(\sim p \vee \sim q)$ is logically equivalent to $p \,\&\, q$, so Jane's beliefs are inconsistent. If this is all we know about Jane's beliefs, we will suspect that her overall state is rationally flawed.

Let me quickly forestall one objection to the setup of this example. One might object that if we heard Jane describe her beliefs that way—especially if she described them immediately one after the other, so she was plainly aware of their potential juxtaposition—we would have to conclude that she uses words like "negation" and "or" to mean something other than our standard truth-functions. Now I would share such a concern about connective meaning if, say, Jane had directly reported believing both "p and q" and "not-q." But we cannot assume that whenever someone has what looks to us like logically inconsistent beliefs it is because she assigns different meanings to logical terms.[51] To do so would be to eliminate the possibility of logical errors, and therefore to eliminate the possibility of a normative theory of (deductive) rational consistency.

There is a delicate tradeoff here. At one extreme, if an apparent logical error is too straightforward and obvious, we look for an explanation in alternate meanings of the connectives. At the other extreme, if what is admittedly a logical inconsistency among beliefs is too nonobvious or obscure, Cognitive Reach concerns may make us hesitant to ascribe rational error. But if we are to have a normative theory of logical consistency at all, there must be some middle zone in which an inconsistency is not so obvious as to impugn connective interpretation while still being obvious enough to count as rationally mistaken. I have chosen a pair of beliefs for Jane that strikes me as falling

[51] At the beginning of my elementary logic course I find students willing to make all sorts of logical mistakes, but I do not interpret them as speaking a different logical language than I.

within that zone. While you may disagree with me about this particular example, as long as you admit the existence of the sweet spot in question I am happy to substitute an alternate example that you think falls within it.

Given what we know of Jane so far, we are apt to return a negative rational evaluation of her overall state. But now suppose we learn that Jane has been taught that this combination of beliefs is rationally acceptable. Jane says to us, "I understand full well that those beliefs are related. I believe that when I have a belief of the form $\sim(\sim x \vee \sim y)$, the only attitude toward y it is rationally permissible for me to adopt while maintaining that belief is a belief in $\sim y$." Perhaps Jane has been led to this belief about rational consistency by a particularly persuasive (though misguided) logic teacher, or perhaps her views about rational consistency are the result of cultural influences on her.[52]

We now have two questions: First, is there any way to fill in the background circumstances such that it's rationally permissible for Jane to have this belief about what's rationally permissible? Second, is there any way to fill in the background circumstances such that Jane's combination of p/q beliefs actually is rationally permissible—such that it's rationally okay for her overall state to contain both a belief in $\sim(\sim p \vee \sim q)$ and a belief in $\sim q$?

I will distinguish three different positions on Jane's case, divided by their "yes" or "no" answers to these two questions.[53] Begin with what I call the "top-down" position, which answers both questions in the affirmative. On this view Jane's training can make it rationally permissible for her to maintain the logically inconsistent beliefs, and also for her to believe that it is rationally acceptable for her to do so. According to the top-down view, Jane's authoritative evidence makes it rationally permissible for her to believe certain belief combinations are acceptable, then that permission "trickles down" to make the combinations themselves permissible as well. One might motivate this position by thinking about the fact that rational requirements are consistency requirements, then concluding that it is the consistency between an agent's attitudes and her beliefs about the rationality of those attitudes that is most important by rationality's lights. On this reading Jane's state need not exhibit any rational flaws.

I will read the top-down position as holding that no matter what particular combination of attitudes an agent possesses, we can always add more to the story (concerning the agent's training, her beliefs about what's rational, etc.) to make her overall state rationally permissible. One could imagine a less extreme top-down position on which certain obvious, straightforward

[52] Again, however we tell our background story about Jane we have to ensure that the connective words coming out of her mouth still mean our standard truth-functions. Perhaps Jane's attitude comes from an authoritative logic professor who taught her the standard truth-functional lore but accidentally wrote the wrong thing on the board one day—a mistake that Jane has unfortunately failed to recognize as such and so has taken to heart.

[53] Technically there are four possible yes-no combinations here, but the view that answers our first question "no" and our second question "yes" is unappealing and I don't know of anyone who defends it. So I'll set it aside going forward.

mistakes are rationally forbidden no matter one's background, then the rational latitude granted by training or testimony grows as mistakes become more difficult to see. To simplify matters I will stick to discussing the pure top-down view, but what I have to say about it will ultimately apply to compromise positions as well. On the pure view no evidence is indefeasible and no combination of attitudes is forbidden absolutely, because an agent could always have higher-order beliefs and evidence that make what looks wrong to us all right.

The opposition to top-down splits into two camps. Both answer our second question in the negative; they split on the answer to the first. What I call the "bottom-up" position holds that it is always rationally forbidden for Jane to believe both $\sim(\sim p \vee \sim q)$ and $\sim q$, and it is also always forbidden for her to believe that that combination is rationally permissible. According to this view, when a particular inference or combination of attitudes is rationally forbidden, there is no way to make it rationally permissible by altering the agent's attitudes about what's rational. What's forbidden is forbidden, an agent's beliefs about what's rational are required to get that correct, and no amount of testimony, training, or putative evidence about what's rational can change what is rationally permitted or what the agent is rationally permitted to believe about it.[54]

Between top-down and bottom-up is a third position, which I call the "mismatch" view. The mismatch view answers our second question "no" but our first question "yes"; it holds that while Jane's education may make it rationally acceptable to believe that her beliefs are permissible, that does not make those beliefs themselves permissible. The mismatch position agrees with bottom-up that Jane's attitudes directly involving p and q are rationally forbidden. But while bottom-up holds that Jane also makes a rational mistake in getting this fact about rationality wrong, mismatch allows that certain circumstances could make Jane's false belief about the rational rationally okay. (For our purposes we need not specify more precisely what kinds of circumstances those are—I'll simply assume that if they exist then Jane's case involves them.) Mismatch differs from top-down by denying that circumstances that rationally permit Jane's believing that her attitudes are acceptable make those attitudes themselves okay.[55]

[54] To be clear, the bottom-up position does not deny the possibility of defeaters in general. For example, if a statistical sample rationally necessitates a particular conclusion it will still be possible for additional, undercutting evidence to reveal that the sample was biased and so change what can be rationally inferred from it. The dispute between top-down and bottom-up views concerns additional evidence that is explicitly about a priori rational requirement truths, and whether such evidence may change both the agent's higher-order beliefs and what's permissible for her at the first order.

[55] Given my insistence on evaluating only overall states—in their entirety—how can we make sense of this talk about the mismatch view's permitting some components of Jane's state while forbidding others? The best way is to think about what overall states in the vicinity the mismatch view takes to be rationally permissible. For example, the mismatch position makes rationally permissible an overall state containing a belief that $\sim(\sim p \vee \sim q)$, a belief that q, and a belief like Jane's that the only attitude toward q permissible in combination with $\sim(\sim p \vee \sim q)$ is

How do the Akratic Principle, the Fixed Point Thesis, and our arguments apply to these positions? Hopefully it's obvious that the mismatch position contradicts the Fixed Point Thesis. On the mismatch reading, it's rationally *impermissible* for Jane to combine a belief in $\sim(\sim p \vee \sim q)$ with a belief in $\sim q$—yet it's *permissible* for Jane to believe this combination of beliefs is okay. Thus the mismatch view would rationally permit Jane to have a false belief about which belief combinations are rationally permissible. As we've seen, the Fixed Point Thesis can be grounded in the Akratic Principle, and the mismatch position is in tension with that principle as well. Mismatch holds that in order for Jane to square herself with all the rational requirements on her, she would have to honor her testimonial evidence by maintaining her beliefs about what's rationally permissible, while at the same time adopting some combination of p/q attitudes like $\sim(\sim p \vee \sim q)$ and q. But then Jane would possess an attitude (or combination of attitudes) that she herself believes is rationally forbidden in her situation, which would violate the Akratic Principle.[56]

The top-down position may also seem to run directly afoul of the Fixed Point Thesis. Absent any cultural or authoritative testimony, it would be rationally forbidden for Jane to believe both $\sim(\sim p \vee \sim q)$ and $\sim q$. Top-down seems to license Jane to believe that that combination of beliefs is permissible, so top-down seems to make it rationally permissible for Jane to have a false belief about what's rational.

Yet the point is a delicate one. The top-down theorist holds that an agent's evidence about what is rationally forbidden or required of her affects what is indeed forbidden or required. On the top-down position, Jane's combination of p/q beliefs would be forbidden on its own, but once her testimonial evidence is added that combination becomes rationally acceptable. Thus the belief Jane forms on the basis of testimony about what's rationally permissible for her *turns out to be true* given that testimony and the belief it generates. Jane's higher-order belief correctly describes the lower-order requirements of rationality on her, so there is no straightforward violation of the Fixed Point Thesis or the Akratic Principle.

Another angle on the same point: Of the three positions we've considered, only mismatch directly contravenes the duality phenomenon I highlighted in Section 1. Both bottom-up and top-down take rational requirements on consistency and inference to stand or fall with requirements on attitudes toward particular propositions. The proposition that Jane's combination of p/q attitudes is rationally permissible is a dual of that permission itself. On the bottom-up reading, both the combination and her belief about that

$\sim q$. Both the top-down position and the bottom-up position would deny that this overall state is rationally permissible.

[56] Acknowledging this tension, Weatherson offers his Kantians argument against the Akratic Principle so he can defend a mismatch position. Ralph Wedgwood's views are also interesting on this front, and have been evolving—Wedgwood (2012) defends a mismatch view, despite the fact that Wedgwood (2007) embraced a version of the Akratic Principle! (Thanks to Ralph Wedgwood for correspondence on this point.)

combination are rationally impermissible. On the top-down reading there are circumstances in which Jane's belief about the combination is rationally permissible, but in those circumstances the combination is permissible as well. Only the mismatch position suggests that Jane could be permitted to believe that a belief combination is required (or permitted) while that combination is in fact forbidden.

So the top-down position does not *directly* conflict with the Fixed Point Thesis in the way mismatch does. Yet I believe that top-down is ultimately inconsistent with that thesis as well. This is because any top-down view is committed to the possibility of an agent's being rationally permitted to believe something false about what's rationally required—if not in her own current situation, then in another. To see why, imagine Jane's case happens in two stages. At first she has no testimony about combinations of p/q beliefs, and simply believes both $\sim(\sim p \vee \sim q)$ and $\sim q$. At this point both bottom-up and top-down agree that her overall state is rationally flawed. Then Jane receives authoritative testimony that this combination of attitudes is rationally permitted, and comes to believe that she is permitted to possess the combination. According to the top-down position, at the later stage this claim about what's permitted is true, and Jane's overall state contains no rational flaws.

But what about Jane's beliefs at the later stage concerning what was rationally permissible for her at the earlier stage? I will argue that according to the top-down theorist, there will be cases in which it's rationally permissible for Jane to believe that at the earlier stage (before she received any authoritative testimony) it was rationally permissible for her to believe both $\sim(\sim p \vee \sim q)$ and $\sim q$. Since that's an a priori falsehood about what rationality requires (even by the top-down theorist's lights), the top-down position violates the Fixed Point Thesis.

Why must the top-down theorist permit Jane such a belief about her earlier situation? One reason is that the top-down view is motivated by the thought that the right kind of upbringing or testimony can make it rational for an agent to believe anything about what's rationally permissible. Suppose the authorities simply came to Jane and told her that believing both $\sim(\sim p \vee \sim q)$ and $\sim q$ was *permissible for her all along*. The top-down view of testimony and its higher-order influence suggests that under the right conditions it could be rational for Jane to believe this.

Even more damning, I think the top-down theorist *has* to take such higher-order beliefs to be permissible for Jane in order to read her story as he does. In our original two-stage version of the story, in which Jane first believes both $\sim(\sim p \vee \sim q)$ and $\sim q$ and then receives testimony making that combination of beliefs rationally permissible, what is the content of that testimony supposed to be? Does the authority figure come to Jane and say, "Look, the combination of beliefs about p and q you have right now is logically inconsistent, and so is rationally impermissible—until, that is, you hear this testimony and believe it, which will make your combination of beliefs rationally okay"? The top-down theorist doesn't imagine Jane's rational indoctrination proceeding via this

sort of mystical bootstrapping. Instead, the top-down theorist imagines that Jane's miseducation about what's rationally permissible (whether it happens in stages or before she forms her fateful p/q beliefs) is a process whereby Jane comes to be misled about what's been rationally permissible for her all along. Even if Jane's beliefs about what's permissible in her own situation are accurate, her beliefs about what's rationally permissible in other situations (including perhaps her own former situation) are false, and are therefore forbidden by the Fixed Point Thesis.

The top-down theorist thinks the right higher-order beliefs can make any attitude combination permissible. But top-down still wants to be a *normative* position, so it has rules for which situational components (such as elements of the agent's evidence) permit which higher-order beliefs. As we saw in the Self-Undermining Argument, these rules come with restrictions to keep from undermining themselves. Once we recognize the possibility of looping, those restrictions broaden to forbid any false belief about what's rational in one's own situation or in others. Properly understood, the top-down position's own strictures make the position untenable.[57] Rational rules form an inviolate core of the theory of rationality; they limit what you can rationally be permitted to believe, even in response to authoritative testimony.

6. PEER DISAGREEMENT

The best objection I know to the Fixed Point Thesis concerns its consequences for peer disagreement. To fix a case before our minds, let's 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 both know that they're peers in this sense. Now suppose that at t_0 Greg and Ben have received and believe the same total evidence E relevant to some proposition h, but neither has considered h and so neither has adopted a doxastic attitude toward 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's and Ben's situation with respect to h is just their total relevant evidence E. Further suppose that for any agent who receives and believes total relevant evidence E, and who adopts an attitude toward h, the only rationally permissible attitude toward h is belief in it. 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

[57] What if we went for a top-down position all the way up—a view on which what's rationally permissible for the agent to believe at higher orders in light of her evidence depends only on what the agent *believes* that evidence permits her to believe, and so on? Such a view would still need normative rules about what counts as correctly applying the agent's beliefs about what's permissible, and those rules could be fed into the Self-Undermining Argument. This point is similar to a common complaint against Quinean belief holism and certain versions of coherentism; even a Quinean web needs rules describing what it is for beliefs at the center to mesh with those in the periphery.

that basis. (To help remember who's who: *G*reg does a *g*ood job rationally speaking, while *B*en does *b*adly.)

At t_1 Greg and Ben have adopted their own attitudes toward h but each is ignorant of the other's attitude. 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 toward h at t_2.

A burgeoning literature in epistemology[58] 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) is receiving renewed attention in moral theory.[59] I'll focus here 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 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.[60] The arithmetic involved will be sufficiently obvious but not too obvious to fall into the "sweet spot" described in the previous section, so Ben's miscalculation is a genuine rational error. While the disagreement literature has certainly not confined itself to entailment cases, as far as I know every player in the debate is willing to accept entailments as a fair test of his or her view.

I will focus primarily on two responses to peer disagreement cases. The Split the Difference view (hereafter SD) holds that Greg, having recognized that an epistemic peer drew the opposite conclusion from him about h, is rationally required to suspend judgment about h.[61] 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.

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 toward h in light of the information that his peer reached a different conclusion from the same evidence. Now what if Ben was an epistemic superior to Greg,

[58] Besides the specific sources I'll mention in what follows, Feldman and Warfield (2010) and Christensen and Lackey (2013) are collections of essays exclusively about peer disagreement.

[59] Of course, discussions of moral disagreement are as old as moral theory itself. The most recent round of discussion includes Setiya (2013), Enoch (2011: Ch. 8), McGrath (2008), Crisp (2007), Sher (2007), Wedgwood (2007: Sect. 11.3), and Shafer-Landau (2003).

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

[61] SD is distinct from the "Equal Weight View" defended by Elga (2007; 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.

someone who Greg knew was much better at accurately completing arithmetic calculations? Surely Greg's opinion about h should budge a bit once he learns that an epistemic superior has judged the evidence differently. Or how about a hundred superiors? Or a thousand? 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 toward h somewhat, *contra* the recommendation of RR.

Call this the Crowdsourcing Argument against RR.[62] 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 whose disagreement with Greg 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 whose disagreement would make it rationally obligatory for Greg 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 he is rationally required to do so. He then meets a hundred experts who believe $\sim h$ on the basis of E. At this point Greg suspends judgment about h. Then he meets another nine hundred experts with the same opinion, and finally caves. Respecting their expertise, he comes to believe $\sim h$.[63]

Once we see the full range of effects the SDer thinks expert testimony can have on Greg, we realize that the SD defender is essentially a top-down theorist. And so his position interacts with the Fixed Point Thesis in exactly the way we saw in the previous section. On the one hand, SD does not produce an immediate, direct violation of the thesis. SD says that at t_2, after Greg meets Ben, the required attitude toward h for Greg is suspension. We stipulated in our case that Greg's original evidence E requires belief in h, but Greg's total evidence at t_2 is now E'—it contains not only E but also evidence about what Ben believes. At t_2 Greg may not only suspend on h but also believe that

[62] You might think of Crowdsourcing 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 Crowdsourcing is an argument *for* anything in particular won't matter in what follows—though we should note that Kelly explicitly endorses the claim I just made that many-superior and single-peer cases lie on a continuum.

[63] The moral disagreement literature repeatedly questions whether there *are* such things as "moral experts" (see e.g. Singer (1972) and McGrath (2008)). If there aren't, this argument may need to be made for practical disagreement cases by piling millions and millions of disagreeing peers upon Greg instead of just one thousand superiors.

suspension is required in his current situation. But since his situation at t_2 contains total evidence E' instead of just E, he doesn't believe anything that contradicts the truths about rationality we stipulated in the case.

Nevertheless, we can create trouble for SD by considering Greg's later-stage beliefs about what was rationally permissible earlier on. If you have the intuition that got Crowdsourcing going to begin with, that intuition should extend to the conclusion that faced with enough opposing experts, Greg could be rationally permitted to believe not only $\sim h$ but also that $\sim h$ was rationally obligatory on E. Why is this conclusion forced upon the SD defender? Again, for two reasons. First, we can stipulate that when the mathematical experts talk to Greg they tell him not only that they believe $\sim h$, but also that they believe $\sim h$ is entailed by E. (It's our example—we can stipulate that if we like!) It would be implausible for the SDer to maintain that Greg must bow to the numerical and mathematical superiority of the arithmetic experts in adopting the outcome of their calculation, but not in forming his beliefs about whether that outcome is correct.

Second, Greg's adopting higher-order beliefs from the experts was probably what the SD defender was envisioning already. When Greg and Ben meet, they have a disagreement not just about whether h is true, but also about whether h was the right thing to conclude from E. SDers often argue that this higher-order disagreement should make Greg doubt whether he performed the calculation correctly (after all, Ben is just as good at figuring these things out as he), and ultimately lead him to suspend judgment on h. Similarly, when the thousand experts come to Greg and convince him to believe $\sim h$, it must be that they do so by telling him his original calculation was wrong. Contrary to what Greg originally thought (they say), E doesn't entail h; instead E entails $\sim h$, so that's what Greg ought to believe. The mathematicians aren't supposed to be experts on the rational influence of testimony; they aren't supposed to be making subtle arguments to Greg about what his total evidence will support after their interaction with him. They're supposed to be telling him something with mathematical content—the type of content to which their expertise is relevant.

And now SD has proved too much: By supposition, E entails h and therefore rationally requires belief in it. When the experts convince Greg that E entails $\sim h$, they thereby convince him that he was required to believe $\sim h$ all along—even before he encountered them. By the Fixed Point Thesis, Greg is now making a rational error in believing that E rationally requires belief in $\sim h$. So it is not rational for Greg to respect the experts in this way. 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.[64]

[64] Exactly how much of the Fixed Point Thesis do we need to get this result? As I see it, all we need is the Special Case Thesis plus the second generalization I described in Section 3. Belief in h is required on E, and after meeting the thousand mathematicians Greg believes that

We now have an argument from the Fixed Point Thesis to the Right Reasons view about peer disagreement. We argued for the Fixed Point Thesis from the Akratic Principle, so if the Akratic Principle is true then misleading evidence at higher levels about what attitudes are required at lower levels does not "trickle down" to permit attitudes that otherwise would have been forbidden. SD and the top-down position both fail because they are trickle-down theories. RR and the bottom-up position 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 said that the best objection to the Fixed Point Thesis comes from its consequences for peer disagreement. Some epistemologists think that on an intuitive basis, Right Reasons (and therefore the Fixed Point Thesis) is simply getting peer disagreement wrong; Ben's general acuity should earn his beliefs more respect, even when he happens to have misjudged the evidence. While we'll return to this thought in Section 7, strictly speaking it isn't an *argument* against RR so much as a straightforward denial of the view. On the other hand, there are now a number of complex philosophical arguments available against RR: that its has deleterious long-term effects, that it leads to illicit epistemic "bootstrapping," etc. I think these arguments have been adequately addressed elsewhere.[65]

Yet there's an objection that immediately occurs to anyone when they first hear RR, an objection that I don't think has been resolved. One can't object to RR on the grounds that it will lead Greg to a conclusion forbidden by his initial evidence; by stipulation the view applies only when he's read that evidence right. But one might ask: How can Greg know that he's the one to whom the view applies—how can he know *he's* the one who got it right? This question may express a concern about guidance, about RR's being a principle an agent could actually apply. Or it may express a concern about Ben: Ben will certainly *think* he got things right initially, so his attempts to respect RR may lead him to

such belief is forbidden. *E* doesn't describe Greg's (entire) situation at the later stage, so we do need that second generalization. But that was the one we were able to establish in Section 4.

The Crowdsourcing continuum also shows another way to argue for the Special Case Thesis's first generalization from Section 3. Suppose we have a view that permits an agent to make rational-requirement errors other than errors in which he takes something to be forbidden that's required (the errors covered by the Special Case Thesis). 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. Facing just Ben, or just the first one hundred experts, didn't compel Greg into any errors covered by the Special Case Thesis (even with its second generalization). But by piling on more experts we could commit the SD defender to the kind of extreme mistake in which an agent inverts what's required and forbidden.

[65] 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.)

form further unsupported beliefs (or at least to resist giving in to Greg when he should).

Here I think 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 the 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.

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 thinks he meets its antecedent, his attempts to follow RR may lead him to adopt the wrong attitude toward 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 t_1 miscalculation of what E rationally requires. It shouldn't surprise us that an agent who initially misunderstands what's rationally required may go on to make further rational mistakes.

Perhaps the objection to RR involves a 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 internalism and the Special Case Thesis.[66] 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 be sufficiently obvious or accessible to both parties at t_1—it lands in our "sweet spot." 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

[66] Once again (see note 39), 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.

determine what RR requires of him are still 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 views.

7. CONCLUSION: ASSESSING THE OPTIONS

This essay began with logical omniscience. Examining formal epistemologists' struggles to remove logical omniscience requirements from their theories, we uncovered a duality phenomenon: 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 toward 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 wherever 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 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 uses only the Akratic Principle as a premise. I then showed that the Fixed Point Thesis has surprising consequences for agents' responses to information about what's rational. If an agent has correctly determined what attitudes her situation requires, rationality forbids changing those attitudes when she receives apparent evidence that she's made the 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 feels preposterous to hold that in the Crowdsourcing case Greg is required to stick to the (admittedly correct) conclusion of his calculations in the face of a thousand acknowledged mathematical experts telling him he's wrong.[67] If this is what the Akratic Principle requires, then perhaps we should drop that principle after all.[68]

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

[68] Thanks to Stew Cohen and Russ Shafer-Landau for discussion of this option.

Unfortunately, dropping the Akratic Principle is no panacea for counterintuitive cases; Horowitz (2013) describes a number of awkward examples confronted by Akratic Principle deniers. Dropping the principle also has difficult dialectical consequences for defenders of Split the Difference (or a compromise like Kelly's Total Evidence View).[69] The mismatch theorist holds that in Crowdsourcing Greg is required to agree with the experts in his higher-order views—that is, he is required to believe along with them that he should believe $\sim h$—but should nevertheless 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 the Akratic Principle. Without the Akratic Principle, Split the Difference is unable to defend itself from the mismatch alternative, on which agents are required to conform their explicit beliefs about what's rational to the views of peers and experts but those beliefs have negligible further effects.

More broadly, I think it's a mistake to assess the Akratic Principle by counting up counterintuitive cases on each side or by treating it and Split the Difference as rational rules on an intuitive par. The Akratic Principle is deeply rooted in our understanding of rational consistency and our understanding of what it *is* for a concept to be normative.[70] Just as part of the content of the concept *bachelor* makes it irrational to believe of a confirmed bachelor that he's married, the normative element in our concept of rationality makes it irrational to believe an attitude is rationally forbidden and still maintain that attitude. The rational failure in each case stems from some attitudes' not being appropriately responsive to the contents of others. This generates the Moore-paradoxicality Feldman notes in defending his Akratic Principle for belief.

While the Akratic Principle therefore runs deep, Split the Difference is grounded in an intuition that can be explained away. I've already suggested that this intuition is a mistaken overgeneralization of the rational significance we assign to testimony in normal situations. And we have a principled explanation for why that generalization gives out where it does. The blank check seemingly written by the rational role of testimony turns out to undermine itself. To maintain rationality's normativity—to enable it to draw *boundaries*—we must restrict rational rules from permitting false beliefs about themselves and each other.

We can also channel some of the intuitive push behind Split the Difference into other, nearby views. For example, we might concede that the mathematical experts' testimony diminishes Greg's *amount* of evidence—or even amount of *justification*—for his conclusion. (Though the Fixed Point Thesis

[69] Christensen (2013) has a particularly good discussion of SD's dependence on the Akratic Principle. See also Weatherson (2013).

[70] A longstanding philosophical tradition questions whether *akrasia* is even *possible*; I am aware of no philosophical tradition questioning whether it's possible to maintain one's views against the advice of experts.

will never allow testimony to swing Greg's total evidence around and all-things-considered support the opposite conclusion.) I would be happy to admit an effect like this in the mirror-image case: If the thousand experts had all told Greg he was absolutely correct, that feels like it would enhance his belief's epistemic status somehow.[71]

If you are convinced that the Akratic Principle should be maintained but just can't shake your Crowdsourcing intuitions, a final option is to hold that Crowdsourcing (and peer disagreement in general) presents a rational dilemma.[72] One might think that in the Crowdsourcing case, Greg's evidence renders rationally flawed any overall state that doesn't concede anything to the experts, while the Fixed Point Thesis draws on Akratic Principle considerations to make rationally flawed any overall state that concedes 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.

Some philosophers deny the existence of rational dilemmas;[73] they will reject this option out of hand. But a more subtle concern is why we went to all this trouble just to conclude that peer disagreement is a rational dilemma. After all, that doesn't tell us what Greg should *do* (or should *believe*) in the situation. We've returned to a concern about the significance of evaluations of rational flawlessness, especially when those evaluations don't straightforwardly issue in prescriptions.

Here I should emphasize again that the evaluations we've been considering are evaluations of *real* agents' overall states, not the states of mythical ideal agents. How can it be significant to learn that such an agent's state is rationally flawed? Consider Jane again, who believes $\sim q$ and $\sim(\sim p \vee \sim q)$ while thinking that belief-combination is rationally permissible. Having rejected the top-down view, we can confirm that Jane's overall state is rationally flawed. While that confirmation 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 immediately tell us what an agent should do in the situation. A number of epistemologists[74] have recently analyzed cases

[71] For a detailed working-out of the justification-levels line, see Eagle (ms). Other alternatives to Split the Difference are available as well. van Wietmarschen (2013), for instance, suggests that while Greg's *propositional* justification for h remains intact in the face of Ben's report, his ability to maintain a *doxastically* justified belief in h may be affected by the peer disagreement.

[72] Although Christensen (2013) employs different arguments than mine (some of which rely on intuitions about cases I'm not willing to concede), he also decides that the Akratic Principle is inconsistent with conciliationist views on peer disagreement. Christensen's conclusion is that peer disagreements create rational dilemmas.

[73] See e.g. Broome (2007).

[74] Such as Elga (ms), Hasan (ms), Weatherson (ms), Schechter (2013), Chalmers (2012: Ch. 2), Christensen (2010), and farther back Foley (1990) and Fumerton (1990).

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.[75] Meanwhile practical philosophers[76] 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, authors then 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 the Akratic Principle, 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 kinds of 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 tradeoffs between unavoidable violations. That's a very different sort of normative task,[77] and the first step in engaging the norms-of-the-second-best involved in sorting out a rational dilemma is to realize that you're in one.[78]

Finally: To conclude that peer disagreements are rational dilemmas is not to deny the Fixed Point Thesis. The thesis holds that no situation rationally permits an overall state containing a priori false beliefs about what situations rationally require. It is consistent with this thesis that in some situations *no* overall state is rationally permissible—in some situations no rationally flawless state is available. So to insist that Greg is in a rational dilemma would not undermine any conclusion I have drawn in this essay.[79] We would still

[75] See, for example, criticisms of Rational Reflection in Christensen (2010) and Elga (2013), and criticisms of single-premise closure in Schechter (2013).

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

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

[78] Imagine someone ultimately develops a robust theory of the second best: some normative notion and set of rules for that notion that determine how one should make tradeoffs and what one should do when caught in a rational dilemma. Will those rules forbid states in which an agent believes the normative notion forbids an attitude yet maintains that attitude anyway? If so, we have a version of the Akratic Principle for that notion, and our arguments begin all over again. . . .

[79] It wouldn't even undermine the Right Reasons position. I have 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 carefully define Split the Difference in a parallel way, then if peer disagreement poses a

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. As long as it's a rational mistake to think or behave as one judges one ought not, it will also be a rational mistake to make false judgments about what's rational.[80]

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.

Audi, R. (1990). Weakness of will and rational action. *Australasian Journal of Philosophy* 68, 270–81.

Balcerak Jackson, M. and B. Balcerak Jackson (2013). Reasoning as a source of justification. *Philosophical Studies* 164, 113–26.

Bergmann, M. (2005). Defeaters and higher-level requirements. *The Philosophical Quarterly* 55, 419–36.

Bjerring, J. C. (2013). Impossible worlds and logical omniscience: An impossibility result. *Synthese* 190, 2505–24.

Brandom, R. B. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.

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

Broome, J. (2007). Wide or narrow scope? *Mind* 116, 359–70.

Brunero, J. (2013). Rational *akrasia*. *Organon F* 20, 546–66.

Chalmers, D. J. (2012). *Constructing the World*. Oxford: Oxford University Press.

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

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

Christensen, D. (2010). Rational reflection. *Philosophical Perspectives* 24, 121–40.

rational dilemma 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. *That* position on peer disagreement contradicts the Fixed Point Thesis, and the Akratic Principle.

[80] For assistance, discussion, and feedback on earlier versions of this essay I am grateful to John Bengson, Selim Berker, J. C. Bjerring, Darren Bradley, Michael Caie, David Christensen, Stewart Cohen, Christian Coons, Daniel Greco, Ali Hasan, Shyam Nair, Ram Neta, Sarah Paul, Miriam Schoenfield, Russ Shafer-Landau, Roy Sorensen, Hank Southgate, Ralph Wedgwood, and Roger White; audiences at Rutgers University, the Massachusetts Institute of Technology, the University of Pittsburgh, Harvard University, Washington University in St. Louis, the University of Arizona, the third St. Louis Annual Conference on Reasons and Rationality, and the fifth annual Midwest Epistemology Workshop; and the students in my Spring 2011 Objectivity of Reasons seminar and my Spring 2012 Epistemology course at the University of Wisconsin-Madison. I am also grateful for two helpful referee reports for *Oxford Studies in Epistemology* from John Broome and John Gibbons.

Christensen, D. (2013). Epistemic modesty defended. In D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays*, pp. 77–97. Oxford: Oxford University Press.

Christensen, D. and J. Lackey (eds.) (2013). *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly* 49, 113–24.

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. 31–9. Oxford: Oxford University Press.

Descartes, R. (1988/1641). Meditations on first philosophy. In *Selected Philosophical Writings*, pp. 73–122. Cambridge: Cambridge University Press. Translated by John Cottingham, Robert Stoothoof, and Dugald Murdoch.

Eagle, A. (ms). The epistemic significance of agreement. Unpublished manuscript.

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. 175–86. Oxford: Oxford University Press.

Elga, A. (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies* 164, 127–39.

Elga, A. (ms). Lucky to be rational. 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*, Volume 1, pp. 69–88. Oxford: Oxford University Press.

Foley, R. (1990). Fumerton's puzzle. *Journal of Philosophical Research* 15, 109–13.

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.

Hegel, G. (1975). *Natural Law*. Philadelphia, PA: University of Pennsylvania Press. Translated by T. M. Knox.

Horowitz, S. (2013). Epistemic akrasia. *Noûs*. Published online first.

Ichikawa, J. and B. Jarvis (2013). *The Rules of Thought*. Oxford: Oxford University Press.

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. 111–74. Oxford: 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. (2013). Rational self-doubt and the failure of closure. *Philosophical Studies* 163, 428–52.

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.

Setiya, K. (2013). *Knowing Right from Wrong*. Oxford: Oxford University Press.

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.

Singer, P. (1972). Moral experts. *Analysis* 32, 115–17.

Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research* 85, 273–300.

Titelbaum, M. G. (2014). How to derive a narrow-scope requirement from wide-scope requirements. *Philosophical Studies*. Published online first.

van Wietmarschen, H. (2013). Peer disagreement, evidence, and well-groundedness. *The Philosophical Review* 122, 395–425.

Weatherson, B. (2013). Disagreements, philosophical and otherwise. In D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays*, pp. 54–76. Oxford: Oxford University Press.

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

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

Wedgwood, R. (2012). Justified inference. *Synthese* 189, 273–95.

Weiner, M. (2007). More on the self-undermining argument. Blog post archived at <http://www.mattweiner.net/blog/archives/000781.html>.

Williams, B. (1981). Internal and external reasons. In *Moral Luck*, pp. 101–13. Cambridge: Cambridge University Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T. (2011). Improbable knowing. In T. Dougherty (ed.), *Evidentialism and its Discontents*, pp. 147–64. Oxford: Oxford University Press.