Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals [1]

Formally-inclined epistemologists often theorize about ideally rational agents--agents who exemplify rational ideals, such as probabilistic coherence, that human beings could never fully realize. This approach can be defended against the well-know worry that abstracting from human cognitive imperfections deprives the approach of interest. But a different worry arises when we ask what an ideal agent should believe about her own cognitive perfection (even an agent who is in fact cognitively perfect might, it would seem, be uncertain of this fact). Consideration of this question reveals an interesting feature of the structure of our epistemic ideals: for agents with limited information, our epistemic ideals turn out to conflict with one another. This suggests that we must revise the way we see ideal agents in epistemic theorizing.

1. Ideal vs. Human-centric Rationality

What would an ideally rational agent believe? Of course, the answer depends on just what kind of ideally rational agent is in question. But when epistemologists consider this question, they don't simply answer "everything true". Rationality, after all, involves reacting correctly to the evidence one has, but does not seem to require having all possible evidence about everything. Thus if we seek to understand rationality by constructing a model of ideally rational belief, we

[1] Thanks to Don Fallis, Hilary Kornblith, Michelle Kosch, Don Loeb, Mark Moyer, Roald Nashi, Derk Pereboom, Tomoji Shogenji, Nico Silins and Peter Sutton for helpful discussions and/or comments on earlier drafts. Special thanks to Louis deRosset and Jonathan Vogel, who each provided extensive and incisive written comments. Versions of this paper were read at NYU, Cornell, UMass, and Brown, and I thank the participants for valuable discussion.

will not concentrate on an omniscient being. Instead, we'll consider a non-omniscient thinker who nevertheless is in certain respects cognitively perfect. We might, for example, stipulate the following kinds of things about such an ideally rational agent's beliefs: They would not be based in wishful thinking. They would be independent of the agent's likes and dislikes. They would respect whatever evidence the agent had. And they would respect the logical relations among claims the agent had beliefs about. Let us call a non-omniscient agent who nevertheless is ideally rational an IRA.[2]

This general approach to theorizing about rationality dovetails nicely with the tradition which relates rationality to thinking logically, and then characterizes rational belief with the aid of formal logic. Those who see belief as a binary, all-or-nothing, kind of state have thus often taken logical consistency and logical closure to be rational ideals. And those who conceive of beliefs as coming in degrees have taken conditions based on probabilistic coherence--which can be seen as little more than applying standard deductive logic to graded beliefs--as ideals.[3]

Of course, this whole formal approach to thinking about rationality has been criticized. The main line of criticism takes off from the fact that ideals such as logical consistency or probabilistic coherence are very clearly far beyond the capacities of any human to achieve--even

[2] For a sense of how widespread this approach is, see the following Stanford Encyclopedia entries: James Joyce on Bayes' Theorem, Sven Ove Hansson on Logic of Belief Revision, James Hawthorne on Inductive logic, Robert Koons on Defeasible Reasoning, and William Talbott on Bayesian Epistemology. Books in this tradition include Savage (1954), Ellis (1979), Horwich (1982), Maher (1993) and Levi (1997).

[3] By "conditions based on probabilistic coherence," I mean not only conditions requiring agents to have precise real-valued degrees of confidence satisfying the laws of probability, but also less restrictive conditions modeling rational degrees of belief by sets of probability functions, or qualitative probabilities. For convenience, I'll use the term "probabilistic coherence" to refer to this whole family of conditions.

more so than complete freedom from prejudice or wishful thinking. Such ideals require, for instance, that an agent believe (or, in the case of coherence conditions, be completely certain of) every logical truth. Why then, it is asked, should rules that might apply to a peculiar sort of imaginary beings--ideal thinkers with limited information--have any bearing on us? The fact that we humans have the particular limitations we do, it is urged, is not just some trivial footnote to epistemology; it's a central aspect of our epistemic predicament. Interesting epistemology--epistemology for humans--must take account of this fact.[4]

I think that this line of criticism should be resisted. While there are certainly some projects in epistemology that must take careful account of human limitations, they do not exhaust interesting epistemology.[5] For example, if one's epistemological project were to characterize our ordinary, casual way of using the words "rational" and "irrational" to apply to people, then it might be hard to see how humanly unattainable ideals would play an important role: everyone fails to live up to humanly unattainable ideals, but we obviously don't call everyone "irrational". But there's little reason to think that epistemology should be restricted to such a thin notion of rationality. (Similarly, ethics should not be restricted to studying moral ideals that are perfectly attained by the ordinary people we'd hesitate to call "immoral.")

A related point applies to the project of developing a notion of rationality that's closely linked to an "ought"-implies-"can" notion of epistemic responsibility. Clearly, we don't want to blame anyone for failing to live up to an unattainable ideal. But there are certainly evaluative

[4] For some representative instances of this line of criticism, see Hacking (1967), Cherniak (1986), Goldman (1986), Kitcher (1992), and Foley (1993).

[5] I cannot make the case for this claim here in full. What follows is a brief sketch, with references to more sustained discussions.

notions that are not subject to “ought”-implies-“can”. I would argue that our ordinary notion of rationality is one of them: when we call a paranoid schizophrenic “irrational”, we in no sense imply that he has the ability to do better.[6]

Another epistemic enterprise in which the importance of highly idealized models might be questioned is the so-called “meliorative project”--epistemology aimed at our cognitive improvement. Some have claimed that any interesting epistemology must be aimed at providing us with guidance to help ourselves (or perhaps others) to think better. I personally doubt that philosophers are particularly well-equipped for this sort of endeavor. But even putting that doubt aside, I see no reason to think that the sole point of epistemology should be the production of manuals for cognitive self-help.[7]

What projects are there, then, which make manifestly unattainable epistemic ideals worth studying? One such project is that of assessing us as a species. After all, there is no reason to suppose--even if we are the cognitive cream of the mammalian crop--that we’re the be-all and end-all of any evaluative epistemic notion we come up with. Indexing epistemic perfection to the cognitive capacities of homo sapiens clearly begs some interesting questions.

But the most important reason for resisting the impatience some express about idealized models of rationality does not depend on the interest of evaluating humans as a species. It is clear that our ordinary rationality judgments are based in assessments of peoples’ levels of performance along certain dimensions of epistemic functioning. And these dimensions may

[6] See Feldman and Conee (1985), Alston (1985) and Christensen (2004, 6.4) for more discussion of this point.

[7] See Christensen (2004, 6.5) for further references and discussion.

well be ones whose extremes are beyond human reach. Freedom from wishful thinking is a plausible example. Predicting consequences of social policies in a way that's untainted by self-interest is another. More examples include evaluating other people's behavior and character without prejudice from emotional ties, or from bigotry based on race or sexual orientation. And a natural candidate for this list is having beliefs that do not violate logic.

If rationality consists (at least partly) in good performance along this sort of dimension, then one natural approach to understanding rationality more clearly is to study candidates for rationality-making qualities by abstracting away from human cognitive limitations, and considering idealized agents who can perfectly exemplify the qualities under consideration. Is logical consistency of all-or-nothing belief a rational desideratum? What about probabilistic coherence of degrees of confidence? How should agents update their beliefs when presented with new evidence? It seems that questions like these may be approached, at least in part, by asking ourselves, "What would an IRA believe?"

Now it is important to see that the suggestion here is not that questions about rational ideals reduce to questions about what ideal agents would believe. Any such reduction would likely run afoul of immediate counterexamples involving, e.g., beliefs about the existence of ideally rational agents.[8] For example, it might be the case that any ideally rational agent would be quite confident that there were conscious beings who could not remember making any cognitive errors; this does nothing to show that such a belief is rationally mandatory in general. But this sort of problem does not, I think, undermine the usefulness of IRAs in studying rationality. It's just that one has to be alert to the distinction between those aspects of an IRA's beliefs which help make it ideally rational, and those that are mere side-effects of the idealization.

It might be insisted that we must still connect considerations about IRAs with claims about us non-ideal agents. However, there are simple, plausible ways of doing this. For example, one attractive thought is that if the constraints that apply to IRAs describe the endpoint of a spectrum, then the closer an actual agent's beliefs are to that end of the spectrum, the better (presumably, ceteris paribus). Efficiency in cars is a nice analogue here: perfect efficiency is impossible, but (ceteris paribus) the more closely one approaches this end, the better. Moral principles also might work this way: I am undoubtedly psychologically incapable of being perfectly fair or generous; but the more closely I approximate perfect fairness and generosity, ceteris paribus, the better.[9]

To my mind, some of the most promising applications of highly idealized theorizing about rationality involve taking probabilistic coherence as a constraint on degrees of belief. Considerations along the lines rehearsed above, I think, show that some of the most common objections to idealizations involving probabilistic coherence, on the grounds that they abstract so far from human limitations, are misguided. I would like, then, to say something like: "Well, of course none of us can be probabilistically coherent, but that's no big deal. We can see that coherence is an ideal in part by showing that IRAs have coherent credences. And as far as my

[8] Williamson (2000, p. 209-210) makes essentially this point.

[9] Zynda (1996) argues along these lines; see also Christensen (2004, ch. 6.5).

own beliefs are concerned, the closer I can come to having coherent credences, the more rational my beliefs will be."[10]

Unfortunately, I now think that the claim that IRAs are coherent is probably false, and that the claim about the rationality of approaching coherence in my own beliefs is at least problematic. The reasons for this are related to, but ultimately quite different from, the worries about idealization described above. They raise what seems to me an interestingly different difficulty for the standard way of using ideal agents in theorizing about rationality, a difficulty flowing from the structure of our epistemic ideals.

2. Ideal Rationality Meets Possible Cognitive Imperfection

The problem I would like to examine involves a very different way in which cognitive imperfection poses an obstacle to taking probabilistic coherence as a rational ideal. The problem arises from an agent's apparently rational reflection on her own beliefs. Let us begin by thinking about a case involving a clearly non-ideal agent:

Suppose I prove a somewhat complex theorem of logic. I've checked the proof several times, and I'm extremely confident about it. Still, it might seem quite reasonable for me to be somewhat less than 100% confident. I should not, for example, bet my house against a nickel that the proof is correct. After all, balancing my checkbook has shown me quite clearly that my going over a demonstrative argument, even repeatedly, is not sure proof against error. Given my thorough checking, my being in error this time may be highly unlikely; nevertheless, it is hard to

[10] This thought presupposes that we can make sense of one's beliefs coming closer to coherence. Zynda (1996) develops a way of making sense of this notion in order to give normative force to the unrealizable ideal of coherence.

deny that I should give it some nonzero credence. Let us call the theorem I've proved T. And let us use M to denote the claim that in believing T, I've come to believe a false claim due to a cognitive mistake. The question now arises: given this sort of doubt, how strongly--ideally speaking--should I believe T?

It seems that my giving some slight credence to M is required by my recognition that I may sometimes exhibit cognitive imperfection. And to the extent that I have any rational credence at all in M, I must have some rational credence in the negation of T (since M obviously entails ~T). So my confidence in T should fall short of absolute certainty; in probabilistic terms, it should be less than 1.

But if something like this is correct, it seems to raise an obstacle to taking coherence as a rational ideal for me--an obstacle quite different from that raised by the fact that coherence is humanly unattainable. For according to this argument, it would not be rational for me to have full confidence in T, a truth of logic. In fact, if I did manage to have the coherence-mandated attitude toward T, the argument would urge me to back away from it. So the problem is not the usual one cited in connection with human cognitive limitations. It's not that I can't achieve the probabilistically correct attitude toward T--in this case, I may well be perfectly capable of that. The problem is that, in the present case, it seems that my beliefs would be worse--less rational--if I were to adopt the attitude toward T that's mandated by probabilistic coherence.

It is worth pointing out that the problem is not just about having maximal belief in logical truths. To see this, suppose I give some positive credence to ~T. Now consider what credence I should give to (~T v C), for some ordinary contingent claim C. If it is different from my credence in C, then my credence in these two contingent claims will violate the principle that logically equivalent claims get equal credence. On the other hand, if my credence in (~T v C) is equal to my credence in C, then I will violate the principle that my credence in a disjunction of logically incompatible disjuncts should be the sum of my credences in the disjuncts.

The basic problem is that coherence puts constraints on my credences based on the logical relations among all the claims in which I have credences--including contingent claims. To the extent that I have doubts about whether certain logical relations hold, and to the extent that those doubts are reflected in my credences, coherence may be violated--even when explicit consideration of logical truths is not involved. For another example, suppose that contingent claim P logically entails contingent claim Q, but I am not absolutely certain of this. In at least some such cases, it would seem that I should then have somewhat higher credence in P than in (P & Q). But if I do, then again I have given logically equivalent contingent claims different levels of credence.

Clearly, this problem should be disconcerting to those of us who would advocate coherence--either the simple version, or one of the standard generalizations--as a component of ideal rationality. To my mind, the threat it poses is significantly deeper than that posed by the fact that probabilistic perfection is not humanly possible. Thus it’s worthwhile seeing whether the one might resist the claim that it would be irrational for me to be coherent.

3. Can I Rationally be Certain of T?

Suppose one were to argue as follows:

Certainty Argument: Granted, I must give ~T at least as much credence as I give to M. But I have the strongest possible kind of justification for full confidence in T--I’ve proved it demonstratively. So I should give it full confidence, and should give ~T, and thus M, zero credence. (After all, my proof of T serves as a proof of not-M!) I may not be a perfect being, but I have the best possible reason for believing T, and thus the best possible reasons for being certain that I haven't come to believe a false claim due to a cognitive mistake.

I think that this argument should not tempt us. To see why, suppose that I work out my proof of T after having coffee with my friend Jocko. Palms sweaty with the excitement of logical progress, I check my work several times, and decide that the proof is good. But then a trusted colleague walks in and tells me that Jocko has been surreptitiously slipping a reason-distorting drug into people's coffee--a drug whose effects include a strong propensity to reasoning errors in 99% of those who have been dosed (1% of the population happen to be immune). He tells me that those who have been impaired do not notice any difficulties with their own cognition--they just make mistakes; indeed, the only change most of them notice is unusually sweaty palms. Here, my reason for doubting my proof, and the truth of T, is much stronger. It seems clear that in the presence of these strong reasons for doubt, it would be highly irrational for me to maintain absolute confidence in T. Yet the certainty argument would, if sound, seem to apply equally to such extreme cases.

Could this verdict possibly be resisted? Could one argue that, initial appearances to the contrary, we actually can embrace the certainty argument, even in the strong doubt case? One way of attempting this would capitalize on distinguishing carefully between two sorts of cases: the bad ones, where the drug has impaired my reasoning and my proof is defective, and the good ones, in which I'm one of the lucky 1% who is immune to the drug's effects and my proof is correct. It might be pointed out that we cannot assume that what would be irrational for the person in the bad case would be irrational for someone in the good case. After all, those in the good case have constructed flawless sound proofs of T, and those in the bad case have made errors in reasoning. To say that what holds for one must hold for the other would be to conflate having a correct proof with seeming to oneself to have a correct proof. So it might be argued that although it would be clearly wrong for most people who find out that they've been dosed to dismiss the resulting doubts, at least if I am in the good case, I am in a different epistemic position, and I may rationally dismiss the doubts.[11]

Now I think that there is something to this point. I would not claim that the epistemic situations of the drug-sensitive person and the immune person are fully symmetrical. After all, the drug-sensitive person in the envisioned type of situation makes a mistake in reasoning even before she finds out about the drug, and the drug-immune person does not. But granting the existence of an asymmetry here does not mean that it is rational for the drug-immune person to disregard the evidence suggesting that he has made an error. And it seems clear--especially when one keeps in mind that those who are affected by the drug don't notice any impairment in their reasoning--that given the evidence suggesting I've made a mistake, it would be irrational for me to maintain full confidence in my reasoning, even if I happen to be in the good case.[12]

Thus we cannot exploit the real epistemic asymmetry between the drug-sensitive and drug-immune people to argue that the latter may after all avail themselves of the certainty argument. And if this is correct, it is hard to see how we can support applying the Certainty

[11] The envisioned argument is inspired by a point Thomas Kelly (2005) makes in a different context, though he should not be saddled with it here.

[12] See Feldman (ms.) and Christensen (forthcoming) for discussion of parallel points relating to the epistemology of disagreement.

Argument even to the original cases involving mild self-doubts raised by memories of misadventures in checkbook balancing. Nothing in the Certainty Argument hinged on the mildness of the doubt about my proof. In fact, it does not seem that even the weak positive reasons for doubt provided by the checkbook-balancing memories are needed to prove the point. Suppose I've never made a mistake in balancing my checkbook or in any other demonstrative reasoning. Surely that doesn't license me in being certain that such mistakes are impossible. And as long as such mistakes are possible, it is hard to see how I can be *certain* that they have not occurred. Even if my reason for doubt is slight, and, so to speak, metaphysical--so slight that in ordinary cases, I wouldn't bother to think about it--still, it would seem irrational to be absolutely certain that I had not come to believe a false claim due to a cognitive mistake. And thus it would seem irrational for me to be absolutely certain of T.

If this is right, it underlies a troubling result for those of us who see coherence as a rational ideal. For the only way I can live up to the ideal of coherence here would seem to be by irrationally dismissing the possibility that a cognitive mistake led me to believe T falsely. Being certain of logical truths seems not only to be something that I can't always do--it seems like something I often shouldn't do. And that makes it hard to see what kind of an epistemic ideal probabilistic coherence could be.

4. Would an Ideally Rational Agent be certain of her own ideality?

The troubling result flows from the fact that I must believe myself to be epistemically fallible. But if rational ideals can be thought of as those that would make an *ideal* *agent's* beliefs rational, perhaps this is not the right way to think about the issue. Perhaps an IRA would not only never make a cognitive error, but would also (rationally) be certain her own cognitive perfection. If that were so, then we could at least hold that an IRA would have probabilistically coherent beliefs. And this might help explain a sense in which coherence was, after all, an epistemic ideal. The idea would be something like this: my self-doubts, which prevent me from rationally being certain of T, are a distracting byproduct of my fallen epistemic state. Consideration of IRAs, who are unaffected by such problems, allows us to see what ideally rational beliefs would be like.[13]

It has been claimed that ideal agents have this sort of self-confidence. Jordan Howard Sobel (1987) argues that what he calls "ideal intellects" not only are probabilistically coherent, but display a number of other features as well: They are always absolutely certain, and correct, about their own credences. They have the sort of trust in their future credences that is embodied in van Fraassen's principle of Reflection. And they are absolutely certain that they are probabilistically coherent. Thus an ideal intellect would not only be absolutely certain of T--she'd also have the sort of high intellectual self-opinion that would seem to be needed to be rationally certain that ~M.

[13] I should note that would not solve the whole problem. We would still need to say something about how ideals that apply to such imaginary agents would relate to rationality assessments for humans. Clearly, this task is complicated if we acknowledge that an ideal for the imaginary agent is one which, at least in some cases, it would be worse for a human agent to approach. I'll return to this issue below.

Sobel defends this conception of an ideal intellect as embodying a kind of full integration and self-possession. He also notes that violation of the ideals can leave an agent open to guaranteed betting losses similar to those that figure in standard Dutch Book arguments. For example, suppose an agent doubts (however slightly) that she's perfectly coherent. If the agent's doubt is realized--that is, if she is actually incoherent--then she is of course susceptible to a classic Dutch Book. But suppose that the agent is actually coherent; she just isn't completely confident that she is. Such an agent will accept a bet in which she will pay the bookie some amount--say $X--if she's coherent, as long as the bookie agrees to pay her enough if she's incoherent. The agent will lose $X on this bet, and the bookie can determine this fact merely by consulting the agent's credences. Thus, as in the standard Dutch Book argument, the bookie can take advantage of the agent by knowing nothing except the agent's credences.[14]

Should we, then, hold that IRAs would have the sort of confidence in their own rational perfection that would preclude the sort of worries that seem to undermine rational certainty in T for human beings? It seems to me that reflection on the motivation behind theorizing about IRAs should make us wary of such a move.

[14] Although Sobel points out the betting vulnerabilities associated with violating his ideals, he sees the main ground for the ideals as lying in our conception of an fully integrated and self-possessed agent. (1997, 72)

As noted above, an IRA, as usually conceived of in theorizing about rationality, is quite different from an omniscient god. The IRA reasons perfectly, and is thus logically omniscient (or at least logically infallible)[15], but the IRA is not assumed to be factually omniscient. This conception of an IRA carries with it no obvious presumption that an IRA would know that she was ideally rational. For such an agent to be rationally confident that she was ideally rational, it would seem that some sort of warrant would be required. But while it seems likely that many IRAs would have excellent evidence of their rational prowess, it also seems unlikely that all of them (or perhaps any of them) could be rationally certain of their own rational ideality.

If an IRA had been around for a long time, and if she had a good memory, she might well have evidence that she possessed an excellent epistemic track record. Unlike most of us, she would never have been corrected for a cognitive error. But it's hard to see how even a very long and distinguished epistemic history could justify the sort of absolute self-confidence at issue here. For it's clearly possible for an agent to think flawlessly up until time t, and then to make a mistake. Clearly, a spotless record up until time t does nothing to tell against this particular possibility.

It's also difficult to see how an agent could be introspectively aware of her own cognitive perfection--or, more precisely, it's hard to see how any sort of introspective awareness could justify absolute self-confidence. Anyone who has experienced some of the common states of consciousness involving diminished cognitive capacities knows that, in some cases, it's pretty

[15] Many use "logically omniscient" to describe the IRA. As Zynda (1996) points out, one might well not want to require full omniscience (i.e., being certain of every logical truth), but rather infallibility (being certain of all logical truths about which the agent has any opinion at all). This distinction will not affect the substance of the discussion below.

easy to tell introspectively that one is epistemically impaired. But not all impairments are evident in this way (and even if they were, there’s no reason to think that all possible impairments would be). So the fact that an agent seems to herself to be thinking with perfect lucidity could hardly justify absolute epistemic self-confidence.

It might also be held that some sort of first-person presumption of rationality must exist which is independent of any reliance on introspection, or on the sort of evidence one might use in making third-person assessments of rationality. Such a presumption might be argued not to need justification by anything else. I do not aim to dispute this sort of claim here. I would only insist that any such presumption would have to fall far short of rendering rational an agent’s absolute confidence that she was absolutely logically inerrant.

What should we think about the argument showing that an agent who doubts her own coherence is vulnerable to guaranteed betting losses? I think that on closer inspection, it turns out to be unpersuasive. Note that in the standard Dutch Book arguments, the bookie offers the agent a set of bets with two properties: (1) the agent finds each bet in the set fair, and (2) the set of bets taken together is logically guaranteed to result in a net loss (for the agent’s side of the bets). The existence of a set of bets with these two properties is the crux of the argument; the imagined bookie adds only entertainment value. In the present argument, if the agent is actually coherent but is not fully confident of this fact, the set of bets the bookie would offer pays the bookie only because it includes a bet which pays him if the agent is coherent. Although the agent would indeed lose money on this set of bets in the actual world--since she is in fact coherent--it is not a set of bets which is logically guaranteed to result in a loss for the agent. The bookie can know that the agent will lose only because the bookie knows the contingent fact that the agent has coherent credences. So it seems that this guaranteed betting loss is more an artefact of the betting situation than an indication of any rational defect on the agent's part.

For all of these reasons, it seems unlikely that an IRA would be rationally certain of her own cognitive perfection. And these considerations also raise an obstacle to arguing that extreme self-confidence flows from a sort of ideal integration or self-possession that characterizes ideally rational intellects. Even if one found this line persuasive in isolation, the considerations above suggest that such extreme self-confidence would be inconsistent with what is clearly central to our conception of an IRA: not having irrationally held beliefs.

5. Can an Ideally Rational Agent be certain of T?

It might be objected, however, that the whole line of argument in the previous section is misdirected. After all, what's directly at issue in our example is just whether the IRA can rationally be certain of ~M. The broader claim discussed above, which concerns the agent's own general rational perfection, is clearly a logically contingent proposition. But ~M follows from T--it's a truth of logic! So the fact that the IRA can't rationally be certain that she never makes logical errors is simply irrelevant. The IRA has solid a priori reason to be certain that in believing that T, she isn't believing a falsehood. No reliance on track records or introspection is required.

Although this argument rightly points out a disanalogy between ~M and general epistemic self-confidence, it seems to me that the disanalogy will not suffice for the use to which the argument would put it. We should first note in general that the fact that an agent has a priori justification for some belief does not render her justification immune to undermining or rebutting by a posteriori considerations. If it did, then that even in the case where Jocko tells me that he drugged my coffee, I would be justified in continuing to believe T. But given that even a priori justifications are vulnerable in this way, it's not clear why the IRA's justification for being absolutely confident in T wouldn't be undermined by any general uncertainty she had about her own cognitive perfection.

We can see this point from a different angle by supposing, as the objection urges, that the IRA may be uncertain of her own logical prowess, while nevertheless being fully certain of both T and ~M. To begin with, let's consider a case in which the IRA, though actually cognitively perfect, doesn't have much confidence at all that she is ultra-reliable when she becomes certain of apparent theorems--let's say she has never checked her theorem-detection by consulting external sources, or even by reconsidering the apparent theorems she has come to believe. Suppose that this moderate self-assessment is rational. Can such an agent nevertheless be rationally certain of T and ~M?

It seems to me unlikely that this will be rational. If the agent can be rationally certain of T and ~M, she presumably can perform similar feats a great many times--there is nothing special about T. So for all of the apparent theorems (say T_1 - $T_{10,000}$) the agent considers, she may rationally be certain of the corresponding propositions ($\sim M_1$ - $\sim M_{10,000}$), each denying that she has mistakenly come to believe a falsehood. Assuming that the agent can keep track of what theorems she has become certain of, she would then seem to have excellent reason to think that she has become certain of 10,000 theorems in a row, without once accepting a false one due to cognitive error. But for the IRA to accomplish this feat without having extraordinary theorem-recognition abilities, something else extraordinary would have to be true. She would have to be extremely lucky (avoiding cognitive errors by sheer luck, or only making cognitive errors that happened not to result in believing false claims), or perhaps be guided by some other force which did have extraordinary powers of theorem-recognition. However, it is hard to see how the fact that an agent is an IRA--the fact that she never does makes a logical mistake--would make it rational for the agent to be at all sure that, insofar as her theorem-recognition abilities might have fallen short, extraordinary luck or guidance resulted in her correctly assessing 10,000 theorems in a row. So it's hard to see why we should think that an IRA could be rationally certain of $\sim M_1$ - $\sim M_{10,000}$ while being only moderately confident in her own theorem-recognition ability.

Could the agent's confidence in T_1 - $T_{10,000}$ make it rational for the agent to have a high degree of confidence that she had extraordinary theorem-proving power? I don't think so. This would be like an agent consulting the gas gauge in her car to determine both the level of fuel and what the gas gauge read, and using the resulting beliefs to rationalize confidence that the gauge was accurate; or looking at a series of colored squares to determine both what color the squares were and how they looked, and using that to make rational her confidence that her color vision was accurate. If the agent begins with a rational moderate degree of confidence in her theorem-recognition abilities, it seems clear that she cannot make higher confidence rational in the manner envisaged.[16]

[16] This argument is adapted from arguments given in a different context by Richard Fumerton (1995, 173ff) Jonathan Vogel (2000) and Stewart Cohen(2002).

I should note that the argument does not presuppose that the IRA could not get any track-record-type evidence of her own reliability. It's just that any such evidence would depend on some way of checking the IRA's proofs. So if others checked the proofs and agreed with them, or if the IRA found theorems she had proved listed in a logic book, or even if the IRA

Would our verdict change if the agent's confidence in her own cognitive perfection were short of certainty, but very high rather than moderate? I think not. For the question is whether the agent can rationally be absolutely certain of the results of her theorem-consideration. Once we see how rational confidence in T is undermined by an agent's moderate views about her own cognitive perfection, it seems clear that even very small doubts about her own cognitive perfection should have some effect in limiting the confidence that it is rational for that agent to have in T.

The problem is just a reflection of the basic fact that lies behind all of the examples we've looked at: that the rationality of first-order beliefs cannot in general be divorced from the rationality of certain second-order beliefs that bear on the epistemic status of those first-order beliefs. This is the reason that, in the case of an ordinary person who has proved a theorem, empirical evidence about being drugged in certain ways can undermine a belief whose justification was purely logical. Thinking about $\sim M_1$ - $\sim M_{10,000}$ is simply a way of amplifying a point that applied to the original ~M: that insofar as an agent is not absolutely confident in her own logical faculties, it is likely to be irrational for her to be absolutely confident in particular beliefs delivered by those faculties.[17]

made multiple attempts to prove the same sentences and got consistent results, that would count for something--after all, there would be some possibility that the IRA could get something other than confirming evidence. The problem with the procedure envisioned is that it completely begs the question of the IRA's theorem-proving accuracy.

[17] This suggests that even if one doesn't harbor doubts about distinguishing logical truths from factual ones, there will still be a sense in which our knowledge of logical truths gets ensnared--via reflection on cognitive fallibility--in the web of belief about ordinary factual matters.

Does this point apply to even the most simple and obviously self-evident-seeming beliefs? If not, there may be a different way to argue that the IRA would have full confidence in ~M_1 - ~$M_{10,000}$. Consider a logical truth that, to us, is maximally obvious--say,

T': Everything is self-identical.

Even if it were granted that we would be irrational to place full confidence in complex logical theorems, it might be claimed that we should at least be able to be absolutely confident in claims such as T'. And if so, we ought to be able to have full rational confidence in the negation of

M': In believing that everything is self-identical, I'm believing a false claim due to a cognitive mistake.

But the IRA, it might well be argued, would experience all logical truths, including T_1 - $T_{10,000}$, as being just as self-evident as T'. So it would, after all, be rational for such a being to be completely certain of ~M_1 - ~$M_{10,000}$. The apparent problem arises only if we're misled by ignoring the IRA's superior ability to see clearly and distinctly in cases where we cannot.

It seems to me that this strategy for supporting the IRA's certainty about ~M_1 - ~$M_{10,000}$ will not work. For even if we grant that all theorems are as simple and obviously self-evident to her as T' is to us, I doubt that the obviousness or self-evidence of T' licences us in being absolutely certain of ~M'. Even if there were some special way of seeing clearly and distinctly that occurs when I contemplate claims like T', I don't think I can rationally be absolutely certain that no drug or demon could make it seem to me that I'm seeing clearly and distinctly when in fact I'm contemplating a falsity. And to the extent that I cannot absolutely preclude that possibility of M', I fall short of rational absolute certainty in T'. For similar reasons, even if all logical truths strike the IRA the way T' strikes me, she cannot absolutely preclude the possibility that her cognitive process have misfired or been interfered with in a way that allows some falsehoods to seem self-evidently true. Thus it seems to me that the IRA cannot rationally be absolutely certain of $\sim M_1$ - $\sim M_{10,000}$, and thus she cannot rationally be absolutely certain of T_1 - $T_{10,00}$.

If the argument of the last two sections is right, then, we are faced with the following sort of problem: given that an agent has the sort of limited evidence IRAs have typically been taken to have, it turns out that there is a tension among three prima facie appealing (though, admittedly, loosely formulated) rational ideals.

1. (LOGIC) An agent's beliefs must respect logic by satisfying (some version of) probabilistic coherence.
2. (EVIDENCE) An agent's beliefs (at least about logically contingent matters) must be proportioned to the agent's evidence.
3. (INTEGRATION) An agent's object-level beliefs must reflect the agent's meta-level beliefs about the reliability of the cognitive processes underlying her object-level beliefs.

The problem we saw was that if a standard IRA satisfied (LOGIC) with respect to her beliefs about theorems, and (EVIDENCE) with respect to her beliefs about the reliability of her own cognitive processes, she could not respect (INTEGRATION) with respect to the connections between these two kinds of beliefs.

There are several different reactions possible here. One could of course take the problem as showing that there's something wrong with at least one of the purported rational ideals--at least, in the ways I've been interpreting them. Since I find each of them quite attractive, though, I'd like to explore two other options. The first is to trace the problem to the peculiarities of the standard kind of IRA that I've been discussing, and to avoid the problem by considering a different kind of ideally rational agent. The second is to develop a revised understanding of the use of ideal agents in theorizing about rationality. I'll discuss these in the next two sections.

6. Can Variant Ideal Agents Avoid the Tension?

If the conflict among the three principles arises only because we are taking our IRA to have incomplete evidence, might we avoid the whole problem by simply dropping this assumption? After all, God, on some standard conceptions, is an agent who is not only perfectly rational, but also perfectly informed. It can be hard to understand how God knows things--it would seem that nothing like our ordinary sources of empirical evidence would be necessary (or, really, of any use at all) for God's omniscience. For my part, I'm not at all sure that it finally makes sense that God could be <u>rationally</u> certain of all truths. But perhaps it does, and if there were such a being, we've seen no reason to think that She would have trouble simultaneously satisfying our three principles.

Now I don't want to explore the tenability of supposing that an omniscient being could rationally be certain of her own rational perfection. For in any case, it seems to me that we cannot simply sidestep our problem by investigating ideal rationality with reference to the beliefs of such a being. A central component of epistemic rationality is having beliefs appropriate to incomplete information. A godlike agent's credences would presumably simply mirror the facts--the agent would be certain of all the truths, have zero confidence in all falsities, and have no intermediate degrees of belief at all.[18] Thus such a model would tell us nothing about a central component of epistemic rationality--the sort of component that's in part captured by something like (EVIDENCE). A useful model of epistemic rationality cannot simply collapse rational belief into truth.

Might there be a non-omniscient ideal agent who could yet be sufficiently free of rational self-doubt to satisfy the three principles? If not, we'd have an argument that rational perfection required factual omniscience. This would, I think, be quite a surprising result. As noted at the outset, rationality seems to be a notion designed in part to abstract from well-informedness. We certainly don't see ordinary cases in which a person lacks information as constituting any sort of lapse in rationality. So it would be surprising that although each of our rational principles seems to be aimed at capturing some aspect of thinking well, and not at some aspect of being well-informed, the three principles together required factual omniscience for their joint satisfaction. On the surface, though, what would be required to satisfy the principles would not be omniscience. We've seen that the agent would need to be absolutely certain that she had not been led by cognitive mishaps to err in believing any of T_1 - $T_{10,000}$, but this does not obviously imply anything about the agent's confidence about, e.g., the number of stars in the Milky Way. So it is not clear to me that only an omniscient being could satisfy the three principles.

Nevertheless, it is also not clear that there is any reasonably neat way of describing an agent whose epistemic powers are less than an omniscient god's, yet who could rationally completely dismiss the sort of doubts about herself that would undermine rational absolute

[18] At least in propositions that have truth values. If, e.g., certain propositions about the future don't have truth values, then even God can't know them.

confidence in the theorems she accepted. Without some clear conception of the epistemic resources such a being would have to have, it's not clear whether she would serve as a useful model for studying principles of rationality. At this point, then, I don't see a way of using a super-knowledgeable variant of the standard IRA to study rational ideals in a context where they don't conflict.

Another way of altering the standard IRA to avoid the tension among the three principles would be to think of an IRA which had no self-doubts because she was completely devoid of beliefs about herself–or, at least, about her own beliefs. After all, it is only when the agent begins to reflect on the possibility of her own epistemic imperfection that the problem seems to arise. Perhaps, instead of imagining an IRA who rationally rejects possibilities of her own error, we could conceive of an IRA who simply never entertains them in the first place.

Again, the question that naturally arises is whether such an agent could be ideally rational. After all, it is not in general rational for an agent to ignore empirical possibilities that bear on the truth of her beliefs. Consider, for example, an ordinary agent who is absolutely certain that it's four o'clock, because her watch reads four o'clock and she hasn't ever considered the possibility that her watch is inaccurate. In this case, it's clear that her absolute confidence betrays a rational failing. Perhaps a closer analogy to our case would be an agent who completely trusted her visual perception, and ignored the possibility that things weren't quite as they appeared. Even an agent who had never seen a mirage would not be rational in having absolute confidence that the world was just the way it looked. We might well think that such an agent was by default entitled to believe that the world was the way it looked, but not that she was entitled to absolute certainty.

Moreover, it’s doubtful that an agent who had no concept of herself having a mistaken belief, or no inclination or capacity to reflect critically on her own beliefs at all, could correctly be categorized as ideally rational. There may be some relatively thin sense of rationality that abstracts away from second-order reflection on an agent’s beliefs.[19] But critical reflection on one’s beliefs is not just something peripheral to rational belief-management--it seems to be a central component of what it is to believe rationally in the fullest sense. And even if we should hesitate to require much in the way of actual second-order reflection, it would seem that if an agent did not reflect on her beliefs at all, and if her beliefs were such that they would be undermined if she did reflect, the agent’s beliefs would not be ideally rational. If that’s right, then it would seem that ideally rational beliefs would be sensitive to second-order considerations of the sort we’ve been discussing. So I don’t think that the tension among our principles can be avoided by positing unselfconscious but ideally rational agents.

I won’t take a stand here on whether the three principles are, in the end, jointly satisfiable, either by an omniscient God or by some lesser being who falls short of complete omniscience. But at this point, I don’t see a way of imagining an idealized agent who satisfies the principles and also can serve as a useful model for studying the question of how non-extreme degrees of belief should be constrained by logical structure. So I’d like to turn now to examine the following question: supposing that there is no useful model of an ideally rational agent who satisfies the three principles, what implications does this have for the study of formal constraints on rationality?

[19] A proposal along somewhat similar lines, applied to knowledge rather than rationality, is made by Ernest Sosa (1997). Sosa distinguishes “animal knowledge,” which does not require

7. Rational Ideals without IRAs

The suggestion that rationality might require violating coherence raises a question about what we should say about the agent--perhaps an imaginary agent with unlimited cognitive powers but limited evidence--who <u>does</u> take self-doubt into account appropriately, and thus violates probabilistic coherence? There seem to be two possibilities. First, one could say that, since such an agent's beliefs would not completely respect logic, the agent would not be ideally rational. On this view, ideal rationality simply could not in general be achieved by an agent who reacted to limited evidence in the best possible way (though perhaps it could be achieved by God). A second option would be to say that, insofar as such an agent achieved the best possible beliefs given her evidence, the agent would be ideally rational. On this view, one would acknowledge that an ideally rational agent might be probabilistically incoherent.

any reflection on an agent's beliefs, from a better kind, "reflective knowledge," which does.

Now I'm not sure that the difference between these two views is much more than verbal. One may see "ideal rationality" as forming the best possible beliefs given one's evidence; or one may see it as perfectly exemplifying all rational ideals. But it is important to see that even if one calls the incoherent agent ideally rational, one is not thereby denying that coherence is a rational ideal. We're quite familiar with other ideals that operate as values to be maximized, yet whose maximization must in certain cases be balanced against, or otherwise constrained by, other values. In scientific theory choice, simplicity and fit with the data are plausible examples of balancing. In ethics, promoting well-being and respecting rights may illustrate a different sort of way in which one ideal constrains another. And tension between ideals has been advocated in epistemology, by those who think we should choose our beliefs (in the all-or-nothing sense) so as to maximize true beliefs while also minimizing false ones.[20] In all of these cases, the fact that ideals can be in tension with one another does not undermine their status as ideals. So we can still see (LOGIC) as a rational ideal once we see how it is to be constrained by (EVIDENCE) and (INTEGRATION).

Because of the particular way in which these three ideals interact, there turns out to be a strange way in which the mere possibility of epistemic misadventure implies an actual epistemic imperfection. The (INTEGRATION)-mandated interaction between our first-order beliefs about logic and our second-order beliefs about ourselves results in something that might be called Murphy's Law for epistemology. The usual version of Murphy's Law states that if it's possible for something to go wrong, it will. The epistemic cousin says that if it's possible that something has gone epistemically wrong (more specifically, if it's possible that I've made a mistake in

[20] Thanks to Don Fallis for reminding me of this example.

thinking about some theorem T), then something has actually gone epistemically wrong (my belief about T falls short of some rational ideal). For either I'm certain of T, in which case my belief fails to reflect appropriately the possibility that I've made a cognitive error, or I'm uncertain about T, in which case my belief fails to respect logic.

What implications does this have for our theorizing about formal conditions on rational belief? If we agree that all the rational ideals cannot be simultaneously realized by a non-omniscient agent, can we still use idealized agents in thinking about how logic should constrain rational belief? If so, will the standard arguments supporting formal conditions on rational belief be affected?

I think that we may continue to use idealized agents in studying formal conditions on rational belief. One way to do this is simply to ignore the fact we've been focusing on: that the standard idealized agent is violating certain strictures about taking self-doubt into account. We could also, more self-consciously, suppose that an agent was cognitively unlimited, in the sense that she could achieve probabilistic coherence, but then stipulate that either (a) she didn't have any second-order beliefs, or (b) she was certain that she was ideally rational, or (c) she didn't take the possibility of her rational imperfection as a reason to be less than fully confident of logical theorems. Having conceived of our agent in any of these ways, we could then consider arguments that such an agent should have probabilistically coherent beliefs. In one of these ways, it seems to me that we could still run standard arguments based on rational constraints on preferences, or based on invulnerability to Dutch Books.

If we do this, we will have to understand what we are doing in a way that departs from the standard way in which people have thought about the idealized agents they've imagined. We can not, in these cases, think of the imaginary agent as ideally rational. For the agent would be irrationally ignoring or rejecting epistemically relevant possibilities, or failing to take them into account rationally in adjusting her beliefs. Nevertheless, the fact that one is not considering the agent as ideally rational does not, I think, undermine the agent's value as a device to help think about a particular dimension of rationality: how logical structure should constrain degrees of belief.

This can be seen by reflecting on the purpose of imagining idealized agents. The purpose of the idealization is in part to abstract away from certain human cognitive limitations, and thus to open up the possibility--which is closed off for agents such as us--of satisfying conditions such as probabilistic coherence. And the idealization should also abstract away from other interfering factors. For example, a Dutch Book argument may assume that the imagined agent values money linearly, and exclusively. The point of this assumption is not that it's particularly rational to value money this way--the purpose is just to isolate one central way in which beliefs and preferences relate to one another. Now if I'm right, it turns out that one thing that can interfere with an agent's beliefs respecting logic completely is the sort of (rational) self-doubt we've been examining. In stipulating away considerations of (even rational) self-doubt, we create a situation in which the logical constraint on belief can be studied in isolation.

It is important to remember that considerations about the beliefs of ideal thinkers should not anyway be thought of as providing a reductive analysis of the concept of a rational ideal. The idea is not that we take a condition to be a rational ideal in virtue of the fact that the condition would be satisfied by an ideally rational agent. So if it turns out that rational ideals are in tension with one another (at least for agents with limited information) we may reasonably allow one rational ideal to be violated in order to study another under limited-information conditions. So the interest of the idealized-agent-based arguments would not be vitiated by acknowledging that the agents involved were not, after all, ideally rational. If coherence can be supported by arguments based on this sort of model agent, that tells in favor of taking it as a rational ideal.

But how could the envisioned sort of ideal have the right sort of evaluative implications for humans? Once we admit that our coherent idealized agent is not actually ideally rational, doesn't the whole exercise lose its epistemic significance?

I think that once we see the structure of epistemic ideals in the way I've been urging, we can see that this is not a problem. It's always been clear that the sort of evaluative principle in question--e.g., the more coherent an agent's beliefs are, the better--must be understood as subject to a ceteris paribus clause rooted in the limitations of an agent's cognitive system. For example, if improving coherence precluded gathering evidence, or required becoming a paranoid schizophrenic, then ceteris wouldn't be paribus, and the agent's beliefs would be less rational if she took the more coherent option.[21] What the above discussion makes clear is that the ceteris paribus conditions must be understood to encompass another dimension. It's not just that our human fleshly limitations might happen to impose epistemic costs on maximizing certain epistemic desiderata. Conflict among epistemic desiderata turns out to flow as well from

[21] Zynda (1996) defends probabilistic coherence as an ideal which imposes prima facie obligations on us.

something much more general: it turns out that our very status as beings with limited information places some epistemic desiderata at odds with others.

So even for us, it still makes sense to say that the more coherent our beliefs are, the better, ceteris paribus. But the ceteris paribus conditions make reference to other epistemic ideals. And if the only way of achieving the probabilistically correct attitude toward some claim T would involve embracing irrational beliefs about my own logical invincibility, or violating the principle that my object-level beliefs should cohere with my meta-level beliefs about the reliability of the cognitive processes behind those object-level beliefs, then adopting the coherent attitude toward T might well render my beliefs less rational.

So: if all this is right, then the tension among epistemic ideals, at least for agents with limited information, requires us to reconceptualize the sorts of ideal agents often considered in studying formal constraints on degrees of belief, but it doesn't undermine their usefulness. And the fact that the ideal of probabilistic coherence may be constrained by other epistemic ideals, and not just by human limitations, doesn't undermine its status as an epistemic ideal.

However, I do worry that other aspects of formal epistemology might not be left undisturbed by the problem I've been discussing. The classic Bayesian view combines a probabilistic coherence requirement with a claim about how beliefs are informed by evidence. Conditionalization, and Jeffrey's generalization of it, are the two standard formal accounts of how evidence bears on belief. Both of these accounts presuppose probabilistic coherence.

One might, of course, study these formal accounts of accommodating evidence by the method I've just recommended for studying formal constraints on an agent's simultaneous beliefs: one might employ probabilistically coherent idealized agents, acknowledging that such agents should not be thought of as ideally rational. And I think that this might well be very useful for studying many cases of evidence bearing on belief. It might even allow us to model cases where some evidential sources undermine others. So a probabilistically coherent ideal agent who employed conditionalization might allow one to model how strongly I should believe that it's four o'clock, given that my watch says it is, and given information about my watch's unreliability.

But I don't yet see how this would allow us to model the way my belief in T should be affected by evidence that Jocko has drugged my coffee. Stipulating probabilistic coherence gives the wrong result: the probability of T, conditional on any evidence at all, will still be 1. The strategy of abstracting away from the conditions imposed by (EVIDENCE) and (INTEGRATION) will not work here, since those conditions are centrally important in determining how evidence about my being drugged affects the level of credence in T it is rational for me to have. So it seems to me that the tension among our epistemic ideals does pose a problem for traditional formal ways of characterizing how evidence bears on rational belief.

I think that this problem might turn out to be difficult to solve, especially formally. This is because the solution would seem to have to respect all three of the principles; and as we've seen, the principles are in some tension with one another. And the correct way of balancing or constraining one ideal by another is likely to prove difficult to capture in a formal system.

I don't want to argue that this problem can't be solved. One might, for example, try the sort of tactic Dan Garber (1983) proposed for handling one version of the old evidence problem. Garber thought the problem stemmed from the assumption of logical omniscience, and so to relax that assumption, he treated certain logical implications metalinguistically, and allowed ideal agents to be less than certain of them. So one might try saying that the credence I should have that T is the probability that the sentence "T" is true, given that I seem to have a proof of it and that I know I have been drugged in a certain way.

This might seem to give the right result in a circumscribed local way. But even this type of model presupposes that the agent is probabilistically coherent over a large range of claims--this is needed for the conditionalization-based mechanism to apply. Garber's particular version assumes that the agent is certain of at least all truth-functional tautologies. But I see no reason to think that proofs of truth-functional tautologies should be exempt from the effects of Jocko's drugs.

Moreover, I suspect that this sort of approach--at least in a simple form--would end up divorcing rational belief too sharply from logic. Even if we restrict our attention to T, and suppose that it's not a truth-functional tautology, the envisioned mechanism would seem to render irrelevant the actual cogency of the agent's reasoning in proving T. Her proof would enter into determining the rationality of her degree of credence in T only as an apparent proof. The fact that T really is a logical truth would have no direct impact on the question of how much rational confidence it merited. Whether certain inferences were logically correct would have no direct impact on the rationality of beliefs supported by those inferences.

The problem with this is especially clear if we think about cases of much milder reasons for doubt. Suppose that Cherry is an excellent reasoner, while Kelly is a poor reasoner, and that the two are separately thinking about some matter. Cherry, through her usual flawless reasoning, becomes highly confident that P. Kelly, through her usual logical blunders, also becomes highly confident that P. It seems to me that we need to count Cherry's confidence in P as more rational than Kelly's. And this remains true even if we add that Cherry's and Kelly's reasons for self-doubt are equivalent (perhaps neither has been given much feedback on her cognitive performance--they both happen to have discovered a few checkbook-balancing errors) and they have the same generally positive assessment of their own reasoning abilities. While the rationality of an agent's belief does depend on the agent's second-order assessment of her reliability, it also depends on other things, including the first-order reasoning on which the belief is based.

Of course, these worries are only preliminary, and it remains to be seen how difficult a problem we're left with. But if the arguments we've been looking at are correct, whatever account we end up giving of the way beliefs should be informed by evidence will have to take into account the interaction among epistemic ideals that we've been examining--in particular, the way that what it's rational for an agent to believe in general is constrained by what it's rational for an agent to believe about herself.

David Christensen

Brown University

REFERENCES

Alston, W. P. (1985), “Concepts of Epistemic Justification,” in his Epistemic Justification: Essays in the Theory of Knowledge (Ithaca: Cornell).

Cherniak, C. (1986), Minimal Rationality (Cambridge: MIT).

Christensen, D. (2004), Putting Logic in its Place: Formal Constraints on Rational Belief (New York: Oxford).

---. (2007), “Epistemology of Disagreement: the Good News,” Philosophical Review 116: 187 - 217.

Cohen, S. (2002), “Basic Knowledge and the Problem of Easy Knowledge,” Philosophy and Phenomenological Research 65: 309 - 329.

Ellis, B. (1979) Rational Belief Systems (Totowa, NJ: Rowman and Littlefield)

Feldman, R. and E. Conee (1985), “Evidentialism,” Philosophical Studies 48: 15 - 34.

Feldman, R. (ms.), “Reasonable Disagreements.”

Foley, R. (1993), Working without a Net (New York: Oxford).

Fumerton, R. (1995) Metaepistemology and Skepticism (Lanham, MD: Rowman and Littlefield).

Goldman, A. (1986), Epistemology and Cognition (Cambridge: Harvard).

Garber, D. (1983), “Old Evidence and Logical Omniscience in Bayesian Confirmation Theory,” in J. Earman, ed., Testing Scientific Theories (Minnesota Studies in the Philosophy of Science 10) (Minneapolis: University of Minnesota Press).

Hacking, I. (1967), “Slightly More Realistic Personal Probability,” Philosophy of Science 34: 311-325.

Horwich, P. (1982) Probability and Evidence (New York: Cambridge).

Kelly, T. (2005), "The Epistemic Significance of Disagreement," Oxford Studies in Epistemology 1, 167 - 196.

Kitcher, P. (1992), "The Naturalists Return," Philosophical Review 101: 53 - 114.

Levi, I. (1997), The Covenant of Reason (New York: Cambridge).

Maher, P. (1993), Betting on Theories (New York: Cambridge).

Savage, L. J. (1954), The Foundations of Statistics (New York: John Wiley & Sons).

Sobel, J. H. (1987), "Self-Doubts and Dutch Strategies," Australasian Journal of Philosophy 65: 56 - 81.

Sosa, E. (1997), "Reflective Knowledge in the Best Circles," Journal of Philosophy 94: 410 - 430.

Vogel, J. (2000), "Reliabilism Leveled," Journal of Philosophy 97: 602 - 623.

Williamson, T. (2000), Knowledge and its Limits (New York: Oxford).

Zynda, L. (1996), "Coherence as an Ideal of Rationality", Synthèse 109: 175 - 216.