

Epistemic Self-Respect¹

David Christensen

Brown University

Everyone's familiar with those annoying types who "think they know everything". Part of what's annoying about many such people is that their self-confidence seems to flow more from the desire to be an expert than from any evidence of actual expertise. Such excessively rosy self-assessments, it seems, indicate something like an epistemic character flaw.

On the other hand, a certain level of epistemic self-respect is plausibly taken to be required by rationality. Consider the extreme examples studied in the literature on epistemic paradox. If I become convinced that I'm a certain kind of incorrigible anti-expert about P--that I'll believe that P if and only if it's not the case that P--then it's hard to see what attitude toward P it could be rational for me to take. If I come to believe P, that will by hypothesis give me reason to think that P isn't true; but if I fail to believe P, that will give me reason to think that P is true. So while scenarios have been devised that would seem to make it rational for me to take myself to be this sort of anti-expert about particular propositions, some (Sorensen 1987) have suggested that I cannot rationally do so (though there is evidently no problem with thinking that others are anti-experts of this extreme sort).

Andy Egan and Adam Elga (2005) have recently argued for more general restrictions on rationally considering oneself an anti-expert. They argue that certain features of rationality require

¹ Thanks to Louis deRosset, Tyler Doggett, Adam Elga, Mark Moyer, Derk Pereboom and the audience at the Aristotelian Society for very helpful discussions or comments on earlier drafts.

that one give a low probability to the claim that one has generally inaccurate beliefs about a given range of propositions. Like Sorensen, they make the connection between taking oneself as an anti-expert and having the sort of beliefs involved in Moore's paradox: "It's raining but I believe that it isn't" would seem to be a particularly clear instance of self-attributing anti-expertise.

They also offer an explanation of why it's unproblematic to take others as anti-experts but problematic in the first-person case. They consider Prof. X, a geologist, who, consulting the track record of Prof. Z, comes to believe that Z is an anti-expert (in the sense of having highly inaccurate beliefs) in geology. This might be perfectly rational on Prof. X's part. However, problems arise when Prof. X finds out that Prof. Z is, in fact, herself. In that case, Egan and Elga write, "she should do something about it. She should change her opinions about geology in such a way that she is no longer an anti-expert. For news of her past anti-expertise is evidence that her present geological views are mistaken" (2005, 86). This seems to indicate that the problem arises not just from anti-expertise being self-attributed, but from its being attributed first-personally.

If rationality precludes taking oneself to be an anti-expert, there is a sense in which a certain kind of epistemic self-respect is built into rationality. I'd like here to examine the idea that rationality entails a kind of respect we must have for our own beliefs that we need not have toward the beliefs of others, and which flows from some aspect of the first-person perspective. I'd like to approach this question from within a framework that takes belief as coming in degrees, and takes probabilistic coherence as a rational constraint on how degrees of belief fit together. To avoid complication, I'll use the simplest form of this framework by taking rational credences to be precise probabilities (though I believe that the lessons I'll arrive at would survive generalization of this simple model).

1. Degrees of Belief and Expertise

When belief is thought of in graded terms, there is a very natural construal that can be given to the notion of expertise: that of a so-called “expert function.” The standard example is of a weatherman. If I take the weatherman to be a complete expert about whether it will rain tomorrow, I’ll treat his beliefs about rain as an excellent rain-indicator. Thus, if I happen to know his degree of credence in rain tomorrow, my credence will simply match his. And even if I’m not sure what his credence in rain is, my beliefs will respect his, in a conditional way: my credence in rain, on the condition that his credence in rain is n , will be n . More formally (where my credences are p , the weatherman’s are q , and R is the proposition that it will rain tomorrow):

$$p(R / q(R)= n) = n.$$

This principle expresses my taking the weatherman as an expert on one particular proposition. But the idea can be generalized to take this sort of principle to hold for large sets of propositions. And certain principles of generalized respect for expert probability functions have been put forward as rational requirements. For example, David Lewis’s (1980) Principal Principle would have a rational agent take objective chances as constraining her credences. And (coming closer to our topic) Bas van Fraassen’s (1984) Reflection principle would have a rational agent treat her own future beliefs with this same sort of respect (here A represents any proposition, p_0 represents the agent’s current probabilities, and p_1 represents her probabilities after some interval of time):

$$\text{(Reflection)} \quad p_0(A / p_1(A)= n) = n.$$

Now Reflection has been criticized. After all, some drug might make one confident that one can fly; if I think I’ll take this drug tomorrow, my present conditional confidence that I’ll be able to fly tomorrow, given that tomorrow I’ll be quite sure I can fly, should not be very high (Christensen

1991). Similarly, I expect to forget that I ate eggs for breakfast this morning. My conditional confidence that I had eggs today--even given that in a year, my confidence in this proposition will be only moderate--is rationally quite high (Talbot 1991).

But let us consider our main question: how should I regard my present beliefs? Is there a way in which I should respect my own beliefs, in the sense of regarding them in a way that precludes anti-expertise? There seems to be a clear disanalogy between the case of distrusting someone else, or my future self, and distrusting my present self. Thinking about belief in a binary way, there is nothing irrational-sounding about my believing that Jocko can't fly, though he thinks that he can, or about my believing that I can't fly, though tomorrow I'll think that I can. But there seems to be something much worse, in a Moore-paradox-type way, about my believing that I can't fly, but that I believe I can.

Thinking about belief in a graded way, it is attractive to think that, even if **Reflection** does not govern rational beliefs about one's future beliefs, its synchronic special case does govern (at least ideally rational) beliefs about one's present beliefs. The synchronic version of **Reflection** looks like this (we can take **SR** to stand for Synchronic Reflection, or perhaps for Self Respect):

$$(SR) \quad p(A / p(A) = n) = n.$$

My credence in A, on the condition that my credence in A is n, is itself n. This does not seem to amount to counting myself as an expert in any obnoxious way--it does not compare me favorably to anyone else. It simply seems to preclude cases in which I disapprove of my own present credences as too strong or too weak. And this sort of disapproval might seem like something that at least an ideally rational agent would not experience.

In fact, a number of philosophers have embraced **SR** as a requirement of reason. Van

Fraassen, while acknowledging that his future-directed Reflection principle has counterintuitive consequences, also refers to “the synchronic--I should think, uncontroversial--part of (Reflection)” (1984, 248). Van Fraassen also ties violations of **SR** to Moore’s paradox (1995, 19). John M. Vickers (2000, 160) refers to **SR** as “a well-known principle of epistemic logic,” and Robert Koons (1992, 23) takes a version of **SR** as a “virtually undeniable principle[] of rationality.”²

It’s worth noting that van Fraassen’s full **Reflection** principle has been supported by a close relative of Dutch Book arguments. These “Dutch Strategy” arguments show that if a bookie knows an Irreflective agent’s credences, he can offer the agent a set of bets which the agent will find fair, but which has the following property: however the agent’s beliefs develop in the future, the bookie will be able to offer further bets, which the agent will then find fair, and which, taken together with the original bets, will guarantee the agent a sure loss. Now these arguments suffer from a weakness that the classic Dutch Book arguments don’t share: they depend on the bookie offering the agent bets at two different times. If the original Dutch Bookie’s profits are best seen as indicators of a kind of inconsistency in the agent’s beliefs, then the Dutch Strategist’s profits are naturally taken as showing inconsistency among the agent’s beliefs at different times. And while synchronic inconsistency is plausibly a rational defect, it is not nearly as plausible that diachronic inconsistency is rationally problematic. (Christensen 1991). However, **SR**, like its diachronic generalization, can also be supported by a Dutch Strategy argument. And unlike the Dutch Strategy argument for forward-looking **Reflection**, the argument for **SR** won’t need to involve bets made at different times. Thus it will escape the charge that it depends on an undefended demand for belief

² See also Skyrms (1980, Appendix 2) and Sobel (1987).

consistency across time.

Finally, **SR** might be motivated more directly along lines parallel to those expressed above by Egan and Elga. Suppose I violate **SR** because I have evidence that I'm optimistic, and I think that my current probability for rain is unrealistically low. One might well ask: if I'm rational, why don't I just raise my credence in rain, correcting for my optimism? If I don't, I seem simply to be failing to take account of evidence--the evidence about my optimistic tendency--that bears on the question about how confident I should be about rain. Put just this way, the idea might seem to involve an unsupportable voluntarism about belief. But a similar thought may be expressed without such implications. If I am ignoring what I myself take to be evidence of my own excessive optimism, then even if I cannot in fact believe otherwise than I actually do, my beliefs seem to fall short of ideal rationality.

To sum up: **SR** seems to have support from a number of directions. Its violation seems to involve ignoring evidence, seems analogous to Moore's-paradoxical thoughts, and leaves violators open to synchronic Dutch Strategies. For all these reasons, **SR** is an attractive candidate for expressing a sort of epistemic self-respect that's required by ideal rationality.

2. An Apparent Counterexample to SR

Despite these attractions, and despite the endorsements of **SR** in the literature, it seems to me that **SR** cannot in the end be a requirement of rationality. The reasons may best be brought out by the following sort of example:

I take it as a truism that people's tendency to believe themselves to be the Messiah is much greater than people's tendency to be the Messiah. Now I don't believe myself to be the Messiah--at

least, not very strongly. Do I give any credence at all to the proposition that I'm the Messiah? Well, I might like to say no. But I suppose it's not impossible that I have some tiny non-0 credence in this proposition. I'd certainly be open to psychological evidence showing that most people brought up in my culture--even those who sincerely disavow it--have some slight tendency toward Messiah complexes, which manifests in part as a small degree of credence that they are the Messiah. So suppose I'm told by some reputable psychologists, who have developed sophisticated ways of measuring credences, that I'm quite likely to have some tiny bit of credence--say, around .0001--that I'm the Messiah. The psychologists may even be able to explain why most people sincerely disavow any belief that they're the Messiah. It seems to me that in this sort of circumstance, it might well be rationally required for me to give their view some credence.

But once I've done this, it's hard to see how I can rationally respect **SR**. After all, the likelihood that I'm the Messiah, even on the supposition that I, like millions of other people, harbor a one-in-10,000 credence in being the Messiah, has got to be less than one in a million:

$$p(M / p(M)= .0001) < .000001$$

So it seems that in these circumstances, at least, a clear violation of **SR** is not only rationally permissible, but rationally required. And if that's so, it furnishes a prima facie counterexample to the claim that **SR** is a general requirement of rationality.

There are, of course, ways of resisting this conclusion. The most tempting, I think, comes from noticing that I would not claim to be ideally rational. So perhaps the fact that I, in certain non-ideal circumstances, should violate **SR** is compatible with taking **SR** as a rational ideal: perhaps the psychologists' information has the effect of interfering with what would, ceteris paribus, be rationally required. This line of thought might be supported by noting that arguments have been

given involving ideally rational agents to show how **SR** is a requirement of ideal rationality. In the next sections, then, I'd like to look at how possible violations of **SR** relate to ideal rationality.

3. Ideal Rationality and Epistemic Self-Respect

One standard way of thinking about rational ideals is to consider the beliefs of an ideally rational agent. Now I think that the relationship between rational ideals and the beliefs of ideally rational agents is not a straightforward matter. But imagining ideally rational agents can serve at least a useful heuristic purpose in theorizing about rationality. So let us begin by asking this: would an ideally rational agent obey **SR**?

Many have held that an ideally rational agent would have probabilistically coherent beliefs. This alone does not imply **SR**. Nevertheless, coherence arguably does imply **SR** given another assumption sometimes made about ideally rational agents: that they have a certain kind of perfect knowledge of their own beliefs (the argument below is an elaboration of one in Sobel 1987). Intuitively, this perfect knowledge means two things: first, the agent is completely confident of what her beliefs are, and second that her self-assessments are accurate.

The first condition is easy to capture, as follows:

Confidence: For any claim A that the agent has some level of credence in, either $p(p(A)=n)=1$ or $p(p(A)=n)=0$.

Either she's certain she has credence n in A , or she's certain that she doesn't.

The second condition is harder to capture in general, since accuracy of credences can't just be measured by truth and falsity. But if we restrict attention to agents who satisfy **Confidence**, the following condition would capture part of what perfect accuracy entails:

Accuracy: if $p(p(A)=n)=1$, then $p(A)=n$; and if $p(p(A)=n)=0$, then $p(A) \neq n$.

This just means that the agent is not wrong in any of her (extremely confident) judgments about her own credences.

Putting the conditions together yields the following sort of self-knowledge: whatever degree of confidence the agent has in A, she's absolutely certain that she has it, and she's absolutely certain that she doesn't have any other degree of confidence in A.

If an agent has this sort of self-knowledge, then we would have a strong argument that she cannot violate **SR** (assuming that she's probabilistically coherent, and assuming the traditional definition of conditional probabilities):

Suppose that the agent has some opinion about her own expertise with respect to A:

$$(1) \quad p(A/p(A)=n) = x.$$

By **Confidence**, she will give probability 0 or 1 to the claim that $p(A)=n$. In the former case, the conditional probability in (1) would be undefined, so it must be that:

$$(2) \quad p(p(A)=n) = 1.$$

By coherence, probabilities conditional on probability-1 claims equal unconditional probabilities, so:

$$(3) \quad p(A/p(A)=n) = p(A).$$

By (2) and **Accuracy**, we have:

$$(4) \quad p(A)=n$$

And putting (3) and (4) together, we get:

$$(5) \quad p(A/p(A)=n) = n.$$

It seems, then, that **SR** flows quite naturally from conditions of self-knowledge which some would

take to characterize the beliefs of an ideal agent. This might be taken, along with the synchronic Dutch Strategy argument, to provide support for taking **SR** as a rational ideal.

4. Self-knowledge, Dutch Strategies and Ideal Rationality

The argument from self-knowledge should prompt the question: are **Confidence** and **Accuracy** really aspects of ideal rationality? On reflection, it seems to me very doubtful that they are.

An initial reason for doubt comes from the sort of consideration adduced by Timothy Williamson (forthcoming). It seems plausible that credences grade into one another insensibly, so that, for example, my being in the condition having 0.2 credence in A is indiscriminable to me from my being in very nearby conditions, such as having 0.2000000001 credence in A. Thus it also seems plausible that if I am in the latter state, it won't be rational for me to be absolutely certain that I'm not in the former state. And if that's so, then when I'm in that latter state, it won't be rational to be absolutely certain that I have the credence that I in fact have.

While I find this reasoning compelling in my own case, I'm somewhat reluctant to rely on it for the case of an ideal agent. One might take rationality to be (at least in part) a matter of a cognitive perfection in belief-management, including monitoring one's beliefs. The monitoring could conceivably require very good access to one's beliefs, and taking that to the maximum might arguably entail perfect access. To put the point another way, an agent's access to her own belief-states might be thought to be part of cognition, and ideal cognition might be argued to be required for ideal rationality. On such a view, the indiscriminability premise that underlies the Williamson-style argument might be resisted. Of course, one might push back here, but I don't want to pursue the issue. For there are independent reasons for denying that an ideally rational agent would have

the sort of self-knowledge described above.

The reasons I have in mind concern not **Accuracy**, but **Confidence**. Even if our ideal agent is granted flawless introspective access to her credences, it is hard to see how she could (rationally) be absolutely confident in this ability of hers. Given that it's clearly possible for agents to make mistakes about their degrees of belief, it's hard to see how a perfectly rational agent could completely dismiss the possibility that she had made this sort of error herself. For there is no reason to think that an ideally rational agent would have any business being absolutely certain that she was ideally rational. The fact that she herself happened to be ideally rational seems like the sort of claim for which some sort of warrant would be needed. But even if the agent had excellent evidence of her own ideality, it's highly implausible that this evidence could rise to the level of rationalizing absolute certainty. Thus it seems to me that an agent who satisfied **Confidence** would thereby fall short of ideal rationality. She would be irrationally confident of her second-order beliefs.³

The upshot of all this is that the sort of ideal self-knowledge which might support **SR** is actually incompatible with ideal rationality. Once an agent takes rational stock of her beliefs about her beliefs, she will not be led thereby to satisfy **SR**.

³ Interestingly, this result about **Confidence** undermines the most natural ways of taking accuracy of second-order beliefs, in a general sense, as a requirement of ideal rationality. Many ways of measuring accuracy quite sensibly increase when (a) one's confidence in true claims increases and (b) when one's confidence in false claims decreases. On such models, accuracy will be maximized only when one has absolute confidence in true claims, and zero confidence in false ones. Thus perfect accuracy requires maximum opinionation. Applied to beliefs about one's own beliefs, this implies **Confidence**. But as we've just seen, **Confidence** is incompatible with ideal rationality. So even if ideal rationality involves, say, having an ideal introspector for one's own beliefs, one cannot rationally place absolute confidence in the deliverances of that introspector.

What about the synchronic Dutch Strategy argument for **SR**? I think that on closer inspection, the kind of betting loss this argument describes should not be seen as indicating any rational defect. Let us consider an example, making the standard assumptions relating my probabilities to my willingness to bet, for an arbitrary violation of **SR**:

Suppose I violate **SR** as follows:

$$(6) \quad p(A | p(A) = .5) = .1$$

Suppose that I give some (perhaps tiny) credence n to $P(A) = .5$ (if conditional probabilities are defined standardly, this must be true if (6) is):

$$(7) \quad p(p(A) = .5) = n$$

The bookie, knowing only my credences, may now offer me the following bets:

Bet 1 pays me $\pounds(1-n)$ if $p(A) = .5$, and pays the bookie $\pounds n$ otherwise. Given (7), I'll consider this bet fair.

Bet 2 is conditional on my credence in A being .5: if it isn't, the bet is off. If my credence is .5, the bet pays the bookie $\pounds 9$ if A, and pays me $\pounds 1$ otherwise. Given (6), I'll consider this bet fair.

Bet 3 will be offered by the bookie only if my credence in A is .5. It will pay me $\pounds 5$ if A, and pay the bookie $\pounds 5$ otherwise. Now, nothing in (6) or (7) guarantees that I'll find this bet fair. But the bookie will only offer it to me on the condition that I have .5 credence in A, in which case I will consider it fair.

Now, consider the possible outcomes. If I don't actually have .5 credence in A, then **Bet 3** is never offered, **Bet 2** is called off, and the bookie takes $\pounds n$ from me on **Bet 1**.

If I do have .5 credence in A, **Bet 3** is offered (and accepted), and **Bet 2** is on. If A is true,

I win $\pounds(1-n)$ on **Bet 1**, lose $\pounds 9$ on **Bet 2**, and win $\pounds 5$ on **Bet 3**, leaving me down $\pounds(3+ n)$. If A is false, I still win $\pounds(1-n)$ on **Bet 1**, and I win $\pounds 1$ on **Bet 2**, but I lose $\pounds 5$ on **Bet 3**, again leaving me down $\pounds(3+ n)$.

Thus the bookie has a strategy for offering bets, each of which I'll take as fair, but which taken together guarantee his profiting at my expense. And he can decide which bets to offer me by knowing nothing over and above my current credences.

Nevertheless, we should notice a difference between these bets and the bets involved in standard Dutch Book arguments. In those arguments, the set of bets is one whose payoffs are logically guaranteed to leave me poorer. Not so for the bets involved in the Dutch Strategy. Consider the case where I don't actually have .5 credence in A. Here, the bookie does not offer me **Bet 3**, so the relevant set of bets is just **1** and **2**. But this set is not one whose payoffs logically guarantee my loss. True, they'll cost me money in the actual world. But in a world where I do have credence .5 in A, and where A is false, I win $\pounds(2-n)$ on these bets. Thus the bookie can know he's going to profit in this case only because he knows that my credence for A is not .5. But that is a contingent fact about the world that, by hypothesis, I don't know. So it's true that the bookie can profit from me by knowing nothing more than my credences. But since the subject matter of the bets includes those very credences, the bookie is profiting by his knowledge of contingent facts beyond my ken. And that sort of guaranteed profit should not be seen as any indication of irrationality on my part.⁴

⁴ A parallel point applies to the case where I have .5 credence in A, and all three bets are in play. The set of all three bets would pay out in my favor in a world where I didn't have .5 credence in A and A was true. Again, the bookie can know that he'll win in the actual world, but
(continued...)

Thus neither of our two ways of supporting **SR** by appealing to conditions on ideal agents--perfect self-knowledge or immunity to guaranteed betting loss--succeeds in the end. In view of the plausibility of the Messiah counterexample, then, **SR** should not be uncontroversial. In fact, it seems quite unlikely that **SR** is a true principle of epistemic logic, or that the sort of epistemic self-respect embodied in **SR** is a general requirement of rationality.

That is not to say, however, that there is nothing to the intuitions that made **SR** attractive in the first place. Some violations of **SR** do seem irrational. In the next section, I'll turn to the question of what we might say about **SR**, even if it isn't a general requirement of rationality.

5. Rationality and Self-Respect

Although we've seen that perfect self-knowledge implies **SR**, we can also see that relaxing the self-knowledge condition just slightly allows clear violations of **SR**. The facts of the Messiah case are compatible with my having quite high confidence in my meta-beliefs, and with those meta-beliefs being highly accurate: Suppose that I give a very low probability to my having .0001 credence in M, and that I actually don't have .0001 credence in M. The fact that I give some (tiny) probability to my having a credence I don't have doesn't violate accuracy dramatically. And we may suppose that I in fact do give a very high probability to my having the credence in M that I actually do have. In this version of the Messiah case, my beliefs need not involve a dramatic departure from perfect self-knowledge. But even a small departure is enough to allow a marked divergence from **SR** that's

⁴(...continued)
only because he knows facts that I don't know concerning the subject matter of our bets.

This problem also gives us an additional reason to reject the diachronic Dutch Strategy arguments discussed above--to my mind, stronger reason than is provided in (Christensen 1991).

also intuitively rational.

What would be required to get violations of **SR** that were intuitively irrational? One clue can be gleaned from comparison with Moore's paradox. Translating a violation of **SR** into rough belief-talk might yield something like this: On the supposition that I believe P, not-P. This seems more like a conditional than a Moorean conjunction; and the two come apart epistemically to the extent that the antecedent is disbelieved. So the fact that, in the Messiah case, I'm quite confident that I don't have the credence in M which I'd take to be unreliable makes the case distinctively un-Moore-like.

This suggests that we look at cases of **SR**-violation where the agent gives high probability to her actually having the belief which she epistemically disrespects. And it seems to me that in such cases, an agent who has reasonably high levels of self-knowledge will end up with irrational beliefs. Consider the following example:

Suppose I'm planning a walk around London tomorrow, and suspect myself of being overly optimistic. I might violate **SR** as follows (where S is the claim that tomorrow will be sunny):

$$(8) \quad p(S / p(S) = .8) = .6$$

And let's suppose that, unlike in the Messiah case, I also give a high probability to my actually having the credence I'd distrust:

$$(9) \quad p(p(S) = .8) = .9$$

Finally, suppose that this meta-belief is accurate--that I do in fact have the level of credence that I'm confident of having:

$$(10) \quad p(S) = .8$$

From (8) - (10), it follows that I'm probabilistically incoherent. Suppressing the arithmetic details,

the reason is roughly this: By (8), I take the claim that $p(S) = .8$ to indicate that the probability of S is .6. By (9), I'm highly confident that $p(S) = .8$. Thus if I am to be coherent, my probability for S must be close to .6 (more precisely, between .54 and .64). But (by (10)) it's not; so I'm incoherent. To put the point even more informally, (8) - (10), taken together, entail that I've failed to take account of a certain bit of evidence in the way I myself think it should be taken account of.⁵

The sunny weather optimism case seems to me to show that there is indeed some truth lying in the region of **SR**. Informally: to the extent I'm confident of having some particular credence c in A, and to the extent that my actual credence in A is close to c , I must come close to satisfying **SR** with respect to my having c in A. (And in the limit, where I'm absolutely certain, and correct,

⁵ As stated, the example employs an unrealistic simplification. Even if I have precise credences, it's still true that for any precise number that I pick--e.g. .8--it is extremely improbable that my level of credence in a proposition such as S is exactly that number. So (9) in the text does not correctly represent the credences I'd have in the imagined situation. And in general, as **SR** is stated, it will virtually never occur that an agent rationally gives a high credence to her having exactly the level of credence mentioned in **SR**'s interior probability.

This suggests that, for **SR** to get a grip in the way illustrated in the example, it really should be generalized. The simplest way would be to replace the equalities with approximate equalities:

$$(\mathbf{SR}') p(A / p(A) \approx n) \approx n.$$

Perhaps better would be a formulation involving ranges of credence:

$$(\mathbf{SR}'') p(A / p(A) \in [m, n]) \in [m, n].$$

Applied to the case in the text, **SR''** would mean, for example, that my credence in S, on the condition that my credence in S is between .7 and .9, must be between .7 and .9.

Even if I can't rationally be very confident that my credence in S is exactly .8, it would seem quite unproblematic for me to be highly confident that my credence is approximately .8, or between .7 and .9. And the argument in the text goes through if we make either set of adjustments in (8) and (9). In the interests of readability, however, I'll continue to suppress this complication in the text. I believe that the substance of the arguments below will not be altered by this simplification. Thanks to Timothy Williamson for pressing this point.

about having *c* in *A*, I must strictly satisfy **SR** with respect to my having *c* in *A*.) Let us call this principle **MSR**, for moderate self-respect.⁶

Should we, then, see **MSR** (or a suitably precisified version thereof) as a principle of epistemic logic expressing a distinctive facet of first-person epistemology--a rational mandate for a certain degree of epistemic self-respect? I have two reservations about giving a positive answer to this question.

First, **MSR** is not really an independent principle of epistemic logic that places special constraints on rational belief. It simply states what follows from coherence under certain conditions. The irrationality of violating **MSR** is nothing over and above the irrationality that's generally involved in breaches of probabilistic coherence.

Now one might acknowledge this, yet hold that since the conditions in which **MSR** constrains belief have to do with the agent's access to her own credences, **MSR** at least expresses a special and distinctively first-person implication of coherence. After all, much of the interest of Moore's paradox has centered around the fact that claims of the form "P but *x* doesn't believe that P" are completely unproblematic unless "*x*" refers in a first-person way. Insofar as there is any

⁶ The argument above provides a view of the situation somewhat different from that offered by Egan and Elga. They assume that rational agents satisfy **Confidence** and probabilistic coherence. They then measure accuracy of graded belief by Brier score, and stipulate that reasonably accurate second-order beliefs are a requirement of rationality. On these assumptions, they show that the degree to which a rational agent may take herself to be an anti-expert (in the sense of having Brier-inaccurate beliefs) is quite limited. They write "It is never rational to count oneself as an anti-expert because doing so must involve either incoherence or poor access to one's own beliefs." (2005, 83) If the arguments above are correct, of course, the assumption of **Confidence** should be abandoned, so Egan and Elga's result would not straightforwardly apply. But it seems to me that the essential insight behind their result is correct. There is a tension among the principles of probabilistic coherence, confidence in one's second-order beliefs, accuracy of one's second-order beliefs, and low epistemic self-respect.

analogy between Moore's paradox and the sort of irrationality that seems to be involved in certain violations of **SR**, it would be surprising if the irrationality did not turn on some particularly first-personal feature of the cases.

However, it seems to me that there is reason to doubt that **MSR**, in the end, expresses even a derivative norm that is deeply first-personal.

We might start by noticing that there is nothing in the above example that turns on whether my **SR**-violating conditional credence involves my representing my own beliefs in a first-person way. It might be that (8) is true in virtue of my thinking, e.g., that DC is overly optimistic about the weather--not even realizing that I am DC. Of course, the ordinary story that would make plausible the strong credence described by (9) would rely on my introspective access to my own beliefs, which would require that I realize that I'm DC. But just as we've acknowledged that one may be wrong about one's own credences, we should acknowledge that one sometimes does have reason to be confident about someone else's credences. Thus there is no inherent barrier to something like (9) being true, even when I'm not thinking of DC in a first-person way.⁷

We can sharpen this point by considering a case which is not in any sense first-person. Suppose that my wife's beliefs about weather--perhaps even unbeknownst to me--happen to be very close to my own. Suppose also that I think that she is too optimistic about tomorrow's weather--in particular, I think the probability of sun, on the condition that she has .8 credence in sun, is only .6 (here I'll use p for my credences, and q for my wife's):

$$(11) \quad p(S / q(S) = .8) = .6$$

⁷ This is especially clear if one thinks in terms of ranges of credences, as opposed to specific credences, as set out in fn. 5 above.

And suppose that I'm confident, and correct, in thinking that my wife does have .8 credence in S.

$$(12) \quad p(q(S) = .8) = .9$$

$$(13) \quad q(S) = .8$$

As in the previous example, the first two suppositions entail that if I'm coherent, my own probability for S must be close to .6. But we are supposing that my credence in S is close to my wife's--i.e., close to .8. So (11), which expresses a violation of an interpersonal Reflection-like principle, makes me incoherent, given our other assumptions about this case. In general, if my credences are close to my wife's (whether or not I know this to be the case), and if I am confident and accurate in my assessments of my wife's credences, then I'm forced by coherence to accord her beliefs a certain kind of epistemic respect.

We're now in a position, I think, to see more clearly the ways in which the epistemic respect we must have for our own beliefs is, and is not, rooted in some aspect or aspects of the first-person perspective. It is of course true that, generally speaking, extreme violations of Reflection-style respect will more often be irrational when directed at one's own beliefs than when directed at the beliefs of others. This asymmetry springs from two sources. First, we are typically in a better position to have strong and accurate opinions about our own credences than to have strong and accurate opinions about the credences of others. Second, our credences are always and automatically maximally close to themselves, but are only sometimes and contingently very close to those of others.

Now the second of these sources does constitute a way in which **MSR** flows from an aspect of first-person epistemology: beliefs are only subject to a demand for the sort of reflective respect we've been examining to the extent that they're close to the agent's own beliefs. This condition

applies automatically to the agent's own beliefs, so it needn't be mentioned in the conditions under which **MSR** applies. A fully impersonal generalization of **MSR** would need to add a clause to the antecedent specifying that the beliefs that were objects of potential respect were close to the agent's beliefs.

Nevertheless, seeing how a requirement of epistemic respect flows from coherence reveals a sense in which it is not at bottom rooted in the first-person perspective. Insofar as we're not confident, or not accurate, in our first-person beliefs about our own beliefs, the demand for **SR**-style epistemic respect simply fails to apply. On the other hand, we may have confident, accurate beliefs about the credences of agents whom we don't think of first-personally, and who may not even be ourselves. Yet, insofar as those agents have credences close to our own, whether or not we realize that they're close, we're rationally required to respect those beliefs, in the sense of not dramatically violating Reflection with respect to them. Clearly, this sort of respect does not depend on the agent's taking any sort of first-person attitude toward the respected beliefs. In fact, the mandate to accord reflective respect to certain beliefs does not even require an agent to have first-person beliefs.⁸

The rational pressure to take Reflective-respectful attitudes toward any beliefs, it seems to me, ultimately flows from the way agents must generally take account of relevance relations among propositions. If an agent's credence in A given E is different from her unconditional credence in A, then there is a sense in which E is, by the agent's own lights, evidentially relevant to A: it is an indicator relevant to A's truth. If her credence in A given E is n, and she's very confident that E

⁸ I believe that very similar points apply to the Egan/Elga results limiting rational self-attribution of inaccurate beliefs.

is true, then her confidence in A will have to be close to n. In the cases in question, E happens to be a claim about someone's credence in A. Suppose that E attributes a credence in A that's close to agent's own. If the agent is very confident in E, then her credence in A conditional on E will have to be close to her simple credence in A. So she can't dramatically violate Reflection with respect to the credence mentioned in E, unless she disregards the way in which E is, by her own lights, an indicator relevant to the truth of A.

Thus I would argue for a claim in epistemology that's parallel to what others have claimed in ethics. It seems to me that the epistemic respect we owe to ourselves is not ultimately different in kind from the epistemic respect we owe to others. In epistemology, however, both varieties of respect flow from a source that is not only impartial, but transcends any special concern with agents as objects of respect. It is the requirement that one not disregard indicators that, by one's own lights, are relevant to the likely truth of the propositions one considers. At bottom, then, epistemic self-respect is just respect for the evidence.

References

- Christensen, David (1991): "Clever Bookies and Coherent Beliefs," Philosophical Review 100: 229-247.
- (forthcoming): "Does Murphy's Law Apply in Epistemology?: Self-Doubts and Rational Ideals," Oxford Studies in Epistemology 2.
- Egan, A., and A. Elga (2005): "I Can't Believe I'm Stupid," Philosophical Perspectives 19(1): 77-93.
- Koons, R. (1992) Paradoxes of Belief and Strategic Rationality (Oxford: Oxford University Press).
- Lewis, D. (1980): "A Subjectivist's Guide to Objective Chance," in his Philosophical Papers vol. II (New York: Oxford University Press, 1986).
- Skyrms, B. (1980): Causal Necessity (New Haven: Yale).
- Sobel, J.H. (1987): "Self-doubts and Dutch Strategies," Australasian Journal of Philosophy 65 (1987): 56-81.
- Sorensen, R. (1987): "Anti-Expertise, Instability, and Rational Choice," Australian Journal of Philosophy 65: 301-315.
- Talbott, W.J. (1991): "Two Principles of Bayesian Epistemology," Philosophical Studies 62: 135-150.
- van Fraassen, B. (1984): "Belief and the Will," Journal of Philosophy 81: 235-256.
- (1995): "Belief and the Problem of Ulysses and the Sirens," Philosophical Studies 77: 7-37.

Vickers, J.M. (2000): "I Believe It, but Soon I'll Not Believe It Any More: Scepticism, Empiricism, and Reflection," Synthèse 124: 155-174.

Williamson, T. (forthcoming) "Why Epistemology Can't be Operationalized," to appear in Q. Smith, ed., Epistemology: New Philosophical Essays (Oxford: Oxford University Press).