Epistemic Self-Doubt

And if I claim to be a wise man,
Well, it surely means that I don't know.
-- *Kansas*

It is possible to direct doubt at oneself over many things. One can doubt one's own motives, or one's competence to drive a car. One can doubt that one is up to the challenge of fighting a serious illness. Epistemic self-doubt is the special case where what we doubt is our ability to achieve an epistemically favorable state, for example, to achieve true beliefs. Given our obvious fallibility, epistemic self-doubt seems a natural thing to engage in, and there is definitely nothing logically problematic about doubting someone else's competence to judge. However when we turn such doubt on ourselves, incoherence seems to threaten because one is using one's judgment to make a negative assessment of one's judgment. Even if this kind of self-doubt can be seen as coherent, there are philosophical challenges concerning how to resolve the inner conflict involved in such a judgment, whether one's initial judgment or one's doubt should win, and why.

Some ways of doubting that we are in a favorable epistemic state are easy to understand and unproblematic. Socrates was confident that he did not know the answers to his most important questions. He believed he did not have the answers, or the right kind of grasp of answers, that would be required for knowledge of what, for example, piety, virtue, and justice are. This recognition led him to avoid endorsing or believing particular answers to his questions, and motivated him to go around town asking others for their answers and making awkward observations about their replies. Though the authorities prosecuted him for this, his offense was not an epistemic irrationality; the belief states he had – doubt about himself that he knew the answers, and lack of confidence in particular answers – fit together sensibly. Moreover, as Socrates told his interlocutors, his acknowledgement that he did not know had the salutary effect of making it possible for him to find out. If he were sure that he already knew, then he would not have motivation to look for the answer.

Not all epistemic self-doubt is so evidently constructive. Socrates could hope to find his answers in the future in part because his doubt was not directed at his faculties for gaining knowledge, and the matters on which he believed himself ignorant were specific and limited. This left him confident of his tools, and still in possession of a lot of knowledge to work with in seeking his answers. For example, it was possible for Socrates to be both sure he did not know what virtue was and yet confident that it was something beneficial to the soul. In contrast, Descartes in his *Meditations* set out to rid himself of all beliefs in order to rebuild his edifice of belief from scratch, so as to avoid all possibility of erroneous foundations. He did so by finding reason to doubt the soundness of his faculty of, for example, sense perception. Instead of casting doubt on his empirical beliefs one by one, he would doubt the reliability of their source and that would blanket all of them with suspicion, loosening the hold that even basic perceptual beliefs had on his mind. Descartes' epistemic self-doubt was extreme in undermining trust in a belief-forming faculty, and in the wide scope of beliefs that were thereby called into question. As in the case of Socrates though, his belief states fit together sensibly; as he convinced himself he might be dreaming, thus undermining his trust that he was in a position to know he had hands, he was also shaken out of his belief that he had hands.

The cases of Socrates and Descartes illustrate that judgments about one's own epistemic state and capacity can provide reasons to adjust one's beliefs about the way things are. Less dramatic cases abound in which rationality's demand for some kind of fit between one's beliefs (first-order beliefs) and one's beliefs about one's beliefs (second-order beliefs) can be seen in the breach. Suppose I am

a medical doctor who has just settled on a diagnosis of embolism for a patient when someone points out to me that I haven't slept in 36 hours. (Christensen 2010a) On reflection I realize that she is right, and if I am rational then I will feel some pressure to believe that my judgment might be impaired, reduce somewhat my confidence in the embolism diagnosis, and re-check my work-up of the case or ask for a colleague's opinion.

Though it seems clear in this case that some reconsideration of the first-order matter is required, it is not immediately clear how strong the authority of the second-order might be compared to the first order in coming to an updated belief about the diagnosis, and there are clear cases where the second order should not prevail. If someone tells me I have unwittingly ingested a hallucinogenic drug, then that imposes some prima facie demand for further thought on my part, but if I know the person is a practical joker and he has a smirk on his face, then it seems permissible not to reconsider my first-order beliefs. There are also cases where it is not obvious which order should prevail. Suppose I'm confident that the murderer is #3 in the line-up because I witnessed the murder at close range. Then I learn of the empirical literature saying that eyewitnesses are generally overconfident, especially when they witnessed the event in a state of stress. (Roush 2009, 252-3) It seems I should doubt my identification, but how can it be justified to throw out my first-order evidence that came from directly seeing that person, in person and close up? An adjudication of some sort between the first order and the higher order is required, but it is not obvious what the general rules might be for determining the outcome of the conflict, or what exactly would justify them.

Questions about epistemic self-doubt can be organized into five over-arching questions: 1) Can the doubting itself, a state of having a belief state and doubting that it is the right one to have, be rational? 2) What is the source of the authority of second-order beliefs? 3) Are there general rules for deciding which level should win the tug of war? If so, what is their justification? 4) What does the matching relation this adjudication is aiming at consist in? 5) If mismatch between the levels can be rational when one first acquires reason to doubt, is it also rationally permitted to remain in a level-splitting state – also known as epistemic akrasia – in which the self-doubting conflict is maintained?

For convenience, approaches to modeling doubt about one's own ability to judge and to the five questions above can be separated into four types, which are overlapping and complementary rather than inconsistent. One approach is through seeing the self-doubting subject as believing epistemically unflattering categorical statements about the relation of her beliefs to the world. Another is through conditional principles, asking what a subject's credence in q should be given that she has a particular credence in q but thinks she may be epistemically inadequate or compromised. A third approach is to construe doubt about one's judgment as a matter of respecting evidence about oneself and one's evidence (higher-order evidence). A fourth approach ties together the first order and seconder order by using the idea that we should match our confidence in p to our expected reliability. That is, treating ourselves like measuring instruments we should aim to be calibrated.

1. How far can consistency and coherence take us? Categorical self-doubting belief

It might seem that consistency and coherence are not strong enough to tell us what the relation should be between the first order and the second order in cases of epistemic self-doubt, in the same way that they don't seem sufficient for explaining what is wrong with Moore-paradoxical statements (crossref SEP, Epistemic Paradoxes). In the latter I assert either "p and I don't believe p", or "p and I believe not-p". There is a lack of fit between my belief and my belief about my belief in either case, but the beliefs I hold simultaneously are not inconsistent in content. What I say of myself would be consistent and quite sensible if said about me by someone else, thus: "p, but she doesn't believe p." Similarly, there is nothing inconsistent in the claim "There is a cat in the distance and she is severely

near-sighted", though there does seem to be a problem with the first-person claim "There is a cat in the distance and I am severely near-sighted" if my assertion about the cat is made on the basis of vision and I give no cue that I mean the second clause as a qualification of the first. My confidence about the cat should have been tempered by my awareness of the limitation of my vision. If there are general principles of rationality that govern self-doubt of our faculties or expertise it seems that they will have to go beyond consistency among beliefs.

However, consistency and coherence do impose constraints on what a subject may believe about the reliability of her beliefs if combined with an assumption that the subject knows what her beliefs are. (This is also found for Moore's paradox, and is used in Shoemaker's approach to that problem. Shoemaker (1994)) One way of formulating an extreme case of believing that one's epistemic system doesn't function well is as attributing to oneself what Sorensen calls anti-expertise. (Sorensen 1988, 392f.) In the simplest type of case, S is an *anti-expert* about p if and only if

Either S believes p and p is false or S does not believe p and p is true. Anti-expertise (A)

Sorensen pointed out that if S is consistent and knows perfectly what her beliefs are, then she cannot believe she is an anti-expert. For if S believes p then, by perfect self-knowledge, she believes that she believes p, but her beliefs that p and that she believes p are together inconsistent with both of the disjuncts of A. Similarly for the case where S does not believe p. This phenomenon generalizes from outright belief to degrees of belief, and from perfect knowledge of one's beliefs to decent but imperfect knowledge of them. (Egan and Elga 2007, 84ff.)

Believing you are an anti-expert is not compatible with coherence and decent knowledge of your own beliefs. Denying that knowledge of our own beliefs is a requirement of rationality would not be helpful, since usefully doubting that one's beliefs are soundly formed would seem to require a good idea of what they are. Egan and Elga favor the view that your response to this fact about anti-expertise should be to maintain coherence and decent self-knowledge of belief, and refrain from believing that you are an anti-expert. However, one can imagine examples where the evidence that you are incompetent is so overwhelming that one might think you should believe you are even if it makes you incoherent. (Conee 1987, Sorensen 1987, 1988, Richter 1990, Christensen 2011)

The problem with self-attributing unreliability coherently doesn't go away if the degree of unreliability is more modest. Consider the following property:

$P(P(q) > .99 \;.\; \neg q \text{ or } P(q) < .01 \;.\; q) > .05$ I'm not perfect (INP)

This says that you are 5% confident that you're highly confident of q though it's false or lack confidence in q though it's true. It is a softened version of anti-expertise, and you can't coherently fulfill it, have $P(q) > .99$, and have perfect knowledge of your beliefs. For in that case $P(P(q) > .99) = 1$, which means INP can only be true if $P(\neg q) > .05$. But $P(\neg q) > .05$ implies $P(q) < .95$ which contradicts $P(q) > .99$. The point survives if you have imperfect but good knowledge of what your beliefs are. The self-doubt expressed through INP is quite modest, but is no more consistent than attributing to oneself anti-expertise, and this will be so for any value of INP's right hand side that is not equal to $P(\neg q)$.

Egan and Elga think that the significance of evidence of anti-reliability is taken into account by seeing it as obligating a subject to revise her first-order belief. However, their view implies that to be rational such a revision must be done without attributing anti-expertise to oneself. One can revise whenever one wants of course, but any revision should have a reason or motivation. If one does not give any credence at all to the possibility one is an anti-expert, then what is one's reason to revise one's first-order belief? There would seem to be no other way to take on board and acknowledge the evidence of your anti-expertise than to give some credence to its possibility. Egan and Elga say that the belief the evidence should lead a subject to is that she has been an anti-expert and that should lead her to revise. (Egan and Elga 2007, 86) But if she avoids incoherence by attributing anti-

expertise only to a previous self, then that belief can't be what leads her to revise her current view. If she does not attribute anti-expertise to her current self then she doesn't give her current self any reason to revise.

The same problem can be seen with Egan and Elga's treatment of cases of self-attributing less extreme unreliability (such as INP) that they regard as unproblematic. Consider a person with growing evidence that his memory isn't what it once was. What effect should this have on his beliefs about the names of students? They compare what happens to his confidence that a given student is named "Sarah" when he hears counterevidence – overhearing someone calling her "Kate" for example – in the case where he has and the case where he hasn't taken the evidence of his decline in memory into account. Via a Bayesian calculation they conclude that when he hasn't taken into account the evidence about his memory, the counterevidence to his particular belief that the student is named Sarah does reduce his belief that she is Sarah, but it does so much less than it would have if he had taken the evidence about his memory into account.

But this analysis represents taking into account the evidence about one's memory only implicitly, as an effect that that evidence already had on one's prior probability that the student is Sarah. That effect is the difference between a .99 and a .90 prior probability, or degree of belief. The distinction that is then derived between the effects that counterevidence can have on the self-doubter and the non-self-doubter is just the familiar point that counterevidence will have a greater effect the lower one's initial probability.

This doesn't tell us how to assimilate news about one's decline in reliability to the first-order belief, but only how to treat other evidence about the first-order matter once one has done so. The question was supposed to be how the evidence about memory should affect our beliefs, and to answer that requires saying how and why that evidence about his memory should make our subject have a .90 rather than .99 initial confidence that the student was Sarah. Surely one must attribute reduced reliability to oneself if one is to have any reason to revise one's first-order belief that the student was Sarah on the basis of evidence of diminished reliability. Even a raw feel of redness of a flower must become an ascription of red to the flower in order for one's experience of it to affect other beliefs such as that it would or wouldn't be an appropriate gift. Even evidence suggesting a small amount of unreliability, as with INP above, presents us with a trilemma: either we incoherently attribute unreliability to ourselves but revise and have a justification for doing so, or we coherently fail to attribute unreliability and revise without justification for doing so, or we remain coherent by failing to attribute unreliability and don't revise, ignoring evidence of our unreliability. It seems that it is not possible for a rational subject to acknowledge evidence of her own unreliability and update her first-order belief on the basis of it.

This approach using consistency (or coherence) plus self-knowledge of belief gives a way of representing what a state of self-doubt is. It sensibly implies that it is irrational to stay in such a state but it also implies that it is irrational to be in it in the first place, making it unclear how self-doubt could be a reason to revise. The approach identifies a kind of matching that rationality demands: give no more credence to the possibility that one is a bad judge of q than one gives to not-q. However, it leaves other questions unanswered. If the rational subject finds herself doubting her judgment, should she defer to her first-order evidence, or her second-order evidence about the reliability of her first-order judgment? What are the rules by which she should decide, and how can they be justified?

2. Conditional Principles

a. Synchronic Reflection and Self-Respect

We might do better at understanding the relations rationality requires between your beliefs and your beliefs about them by adding to the requirements of consistency and coherence a bridge

principle between the two orders expressed via conditional (subjective) probability. Conditional probabilities say what your degree of belief in one proposition is (should be) given another proposition, here the relevant propositions being a first-order proposition q, and the proposition that one has degree of belief x in q, respectively. A first pass at how to represent a situation where my beliefs at the two levels don't match comes from its apparent conflict with the synchronic instance of the Reflection Principle (van Fraassen 1984).

$P_0(q/P_1(q) = x) = x$ Reflection

Reflection says that my current self's degree of belief in q given that my future self will believe it to degree x should be x. It is implied by the fact that her degrees of belief are represented as probabilities that my future self is coherent, but that alone does not rule out the possibility that her judgment is compromised in some other way – as for example when Ulysses anticipated that he would be entranced by the Sirens – and the principle can be questioned for such cases. (Sobel 1987, Christensen 1991, van Fraassen 1995) However, the self-doubt we are imagining is one that the subject has about her current beliefs, and the synchronic version of Reflection

$P_0(q/P_0(q)=x) = x$ Synchronic Reflection (SR)

which says that my degree of belief in q now given that I now believe q to degree x should be x, seems less open to question. Christensen 2007b also calls this principle Self-Respect (SR). This is not the tautology that if I believe q to degree x then I believe q to degree x, for in a logically equivalent form the principle is

$[P_0(q \,\&\, P_0(q)=x) \mid P_0(P_0(q)=x)] = x$ Synchronic Reflection/Self-Respect (SR)

which does not follow from either deductive logic or the probability axioms alone. But SR has been widely endorsed as unobjectionable, and according to some even undeniable, as a requirement of rationality. (van Fraassen 1984, 248; Vickers 2000, 160; Koons 1992, 23; Skyrms 1980 sees as useful a version of it he calls Miller's Principle, though he also shows it is subject to counterexamples.)

While I can sensibly imagine my future self to be epistemically compromised, unworthy of my deference, violating SR would require regarding my current self as epistemically compromised, as having a degree of belief that should be other than it is. This appears to be something that doubt of my own judgment would call for, in which case whether self-doubt can be rational depends on whether SR is a requirement of rationality.

SR can be defended as a rational ideal by Dutch Strategy arguments, though not by the strongest kind of Dutch Book argument. (Sobel 1987; Christensen 1991, 2007b, 328-330, 2010b; Briggs 2010; Roush 2016 argues it can't be defended as a requirement by a Dutch book argument at all.) It has been argued to be questionable if not false on grounds that it conflicts with the *Epistemic Impartiality Principle* that says we should not in general take the mere fact that we have a belief as a reason to have that belief any more so than we do with the mere fact that others have that belief. (Christensen 2000, 363-4, Evnine 2008, 139-143; Roush 2016)[1]

Nevertheless the probabilist – one who thinks that rationality requires probabilistic coherence – will have a difficult time resisting SR since, analogously to what we saw above with anti-expertise, SR follows from coherence if it is supplemented with the further assumption that the subject has perfect knowledge of her own beliefs. Still, this does little to explain intuitively why SR should be binding; even someone who has perfect knowledge that he has a belief should be able to sensibly wonder whether it is a belief he ought to have. That perfect knowledge of our beliefs is a

[1] As a conditional probability SR would also underwrite by conditionalization the mere having of a belief as a reason to continue to have that belief, a view called *epistemic conservatism*, which some have found objectionable. See Sklar 1975, Foley 1982, Adler 1990, and Christensen 1994 for discussion.

requirement of rationality can be doubted in a variety of ways (Williamson 2000; Christensen 2007b, 327-328; Roush 2016). However, as above so for the discussion here, denying that rationality requires knowledge of our own beliefs does not overcome the problem. Self-correction that will be of any use requires some degree of accuracy about one's beliefs, and even if a subject doesn't have perfect self-knowledge coherence still makes reflective demands; Christensen (2007b, 332) has noted that the closer a coherent subject comes to perfect knowledge of his beliefs the more nearly he will satisfy SR.

SR has something to recommend it, but it seems to be a rule a self-doubter will violate. Consider our underslept doctor. It looks as if once it is pointed out to her how long it has been since she slept, she should regard her current confidence in q, her diagnosis of embolism, as higher than it ought to be. I.e., she would instantiate a principle we could call Refraction:

$P_0(q/P_0(q)=x) < x$ Refraction

Apparently, her degree of belief that it's an embolism given that she has degree of belief x that it's an embolism should be less than x, contradicting SR. Or imagine that the person who tells me a hallucinogenic drug has been slipped into my coffee is a trusted friend who is neither in the habit of joking nor currently wearing a smirk on his face. I seem to have an obligation to regard some of my current degrees of belief as higher than they should be.

Refraction is a way of representing a state of self-doubt, one in which I don't regard the degree of belief I (think I) have as the right one to have. But despite the fact that it doesn't have the subject attributing unreliability to herself categorically as we had in the last section, Refraction is not compatible with the combination of coherence and knowledge of one's beliefs, since the latter two together imply SR. In this representation of what self-doubt is, it is not rational according to the probabilistic standard.

One might defend this verdict by saying that the exception proves the rule: if I really think that my degree of belief in q should be different than it is, say because I realize I am severely underslept, then surely I should change it accordingly until I come to a credence I do approve of, at which point I will satisfy SR. However even if it is ideal to be in the state of self-respect that SR describes, it seems wrong to say that a state of disapproving of one's first-order belief when faced with evidence of one's impaired judgment is irrational. In such a case it would seem to be irrational *not* to be in a state of self-doubt.

If we represent epistemic self-doubt as a violation of Synchronic Reflection (Self-Respect), then it is not rational for a coherent person who knows what her beliefs are. This is a general rule giving the same verdict in all cases, that the orders must match, and it gives the form of that matching in terms of a conditional probability. The second-order is in the driving seat since the condition in SR's conditional probability that determines a value for the first-order proposition q is itself a statement of probability, but SR can't lead to change of first-order belief unless one does not know what one's belief in q is. As in the approach via categorical statements above, this simple approach via conditional probability does not represent the cycle of self-doubt and resolution as available to a rational subject.

b. What would the maximally rational subject do?

Another way of representing self-doubt using conditional probability follows the intuitive thought that I should tailor my first-order degree of belief to the confidence I think the maximally rational subject would have if she were in my situation. (Christensen 2010b, 121) This would be a sensible explanation of the authority of higher order beliefs, of why taking them into account would be justified. It gives half of an answer to the question when the first-order should and shouldn't defer to

the second order by identifying a class of second-order statements to which the first order must always defer. However it pushes back the question of which individual statements those are to the question of which probability function is the maximally rational.

A conditional principle that would capture the idea of deferring to the view of an ideal agent who was in one's shoes is:

$Cr(q/P_M(q) = x) = x$

(Christensen 2010b) which says that one's credence in q given that the maximally rational subject in one's situation has credence x in q, should be x. The maximally rational subject obeys the probability axioms and possibly has further rationality properties that one might not possess oneself, though she is assumed to be in your situation, having no more evidence than you do. If one obeys the probability axioms oneself, then this principle becomes:

$P(q/P_M(q) = x) = x$ RatRef

This says that your credence in q should be whatever you take the maximally rational credence to be for your situation, an idea that seems hard to argue with. It is a variant of a principle used by Haim Gaifman (1986) to construct a theory of higher-order probability. There the role of P_M was given to what has become known as an *expert function* corresponding, in his use, to the probabilities of a subject who has maximal knowledge.

RatRef gives a sensible account of cases like the underslept doctor. It would say the reason that the realization she is sleep-deprived should make her less confident of her diagnosis is that a maximally rational person in her situation would have a lower confidence. Moreover this provides us with a way of representing the state of self-doubt coherently, even with perfect knowledge of what one's belief states are. One can have degree of belief y in q, and one can even believe that one has degree of belief y, that is, believe that P(q) = y, consistently with believing that the maximally rational agent has degree of belief x, i.e., $P_M(q) =x$, because these are two different probability functions.

Because self-doubt is not defined as a violation of the conditional probability RatRef, as it was with SR, we can also see a revision that takes you from self-doubt to having your confidence match that of the maximally rational subject as rational according to conditionalization. You may have degree of belief y in q, discover that the maximally rational subject has degree of belief x, and because you have the conditional probability RatRef put yourself in line with that ideal subject. Note that it is not necessary to have an explicit belief about what your own degree of belief in q is for this revision to occur or to be rational.

RatRef has problems that are easiest to see by considering a generalization of it:

$P(q/P' \text{ is ideal}) = P'(q)$ Rational Reflection (RR)

Rational Reflection (Elga 2013) maintains the idea that my degree of belief in q should be in line with that I think the maximally rational subject would have in my situation, but also highlights the fact that my determining what this value is depends on my identifying which probability function is the maximally rational one to have. I can be coherent while being uncertain about that, and there are cases where that seems like the most rational option. This by itself is not a problem because RR is consistent with using an expected value for the ideal subject, a weighted average of the values for q of the subjects I think might be the maximally rational subject. But it is not only I who may be uncertain about who the maximally rational subject is. Arguably the maximally rational subject herself may be uncertain that she is – after all, this is a contingent fact, and one might think that anyone's confidence in it should depend on empirical evidence. (Elga 2013)

The possibility of this combination of things leads to a problem for RR, for if the subject who is actually the maximally rational one is unsure that she is, then if she follows RR she won't fully trust

her own first-order verdict on q but will as I do correct it to a weighted average of the verdicts of those subjects she thinks might be the maximally rational one. In this case my degree of belief in q *given* that she is the maximally rational subject should not be her degree of belief in q. It should be the one she would have in case she were certain that she were the maximally rational subject:

P(q/P' is ideal) = P'(q/P' is ideal) New Rational Reflection (NRR)

This principle (Elga 2013) also faces problems, which are developed below through the approach to self-doubt via higher-order evidence.

The approach asking what the maximally rational subject would do gives a motivation for the idea that second-order evidence has authority with respect to first-order beliefs, and like the SR approach puts the second order in the driver's seat by having a statement of probability in the condition of the conditional probability. This may appear to give unconditional authority to second-order evidence, but second-order evidence won't change the subject's first-order verdict if she believes the latter is already what the maximally rational subject would think. The approach represents self-doubt as a coherent state that one can also coherently revise via conditionalization. It identifies a state of matching between the orders – matching one's confidence to one's best guess of the maximally rational subject's confidence. This gives a general rule, and demands that same matching for all cases, though gives no explicit guidance about how to determine which is the maximally rational subject or degree of belief.

3. Higher-order Evidence

Questions about the rationality (or reasonableness or justifiedness) and import of epistemic self-doubt can be developed as questions about whether and how to respect evidence about one's evidence. Higher-order evidence is evidence about what evidence one possesses or what conclusions one's evidence supports. (crossref SEP Evidence) This issue about the upshot of higher-order evidence does not in the first instance depend on whether we take such evidence as necessary for justification of first-order beliefs. The question is how our beliefs should relate to our beliefs about our beliefs when we do happen to have evidence about our evidence, as we often do. (Feldman 2005, Christensen 2010a, Kelly 2005, Kelly 2010)

Self-doubt is a special case of responding to higher-order evidence. Not all evidence about our evidence arises from self-doubt because not all such evidence is about oneself, as we will see below. Also, representing self-doubting situations as responding to evidence about my evidence takes information about my capacities to be important just insofar as it provides evidence that I have either incorrectly identified my evidence or incorrectly evaluated the support relation between my evidence and my conclusion. For example, in the case of the doctor above who receives evidence that she is severely underslept, the reason she should reconsider her diagnosis is because this is evidence that she might be wrong either in reading the lab tests or in thinking that the evidence of lab tests and symptoms supports her diagnosis. By contrast the fourth approach below via calibration does not see the implications of self-doubt as necessarily proceeding via evidence about our evidence or evidential support.

Like the approach via the maximally rational agent, the evidential approach has the virtue of identifying a justification for responding to the second-order beliefs that self-doubt brings. Their authority comes from the facts that they are evidence relevant to whether one has good evidence for one's first-order belief, and that one should respect one's evidence. This raises the hope that what we already know about evidence can help settle when negative second-order evidence should override a first order belief and when not. Many authors have thought that in either kind of case rationality demands that the two orders eventually match, in some sense, but we will see below that more recent thinking about evidence has led some to defend the rationality of having the first- and

second-order beliefs in tension, for some cases. Another virtue of the evidential approach is that merely knowing what your beliefs are doesn't automatically imply that a state of self-doubt is inconsistent or incoherent as it did in first two approaches above, via categorical beliefs and conditional principles. There is no obvious contradiction in believing both q and that one's evidence doesn't support q, even if one also has a correct belief that one believes q, so what may be irrational about the state must be based on further considerations.

The higher-order evidence approach can be usefully developed through the example of hypoxia, a condition of impaired judgement that is caused by lack of sufficient oxygen, and that is rarely recognized by the sufferer at its initial onset. Hypoxia is a risk at altitudes of 10,000 feet and higher. (Christensen 2010b, 126-127) Suppose you are a pilot who does a recalculation while flying, to conclude that you have more than enough fuel to get to an airport fifty miles further than the one in your initial plan. Suppose you then glance at the altimeter to see that you're at 10,500 feet and remember the phenomenon of hypoxia and its insidious onset. You now have evidence that you might have hypoxia and therefore might have misidentified the support relations between your evidence and your conclusion. Are you now justified in believing that you can get to the more distant airport? Are you justified in believing that your evidence supports that claim?

Letting F be the proposition that you have sufficient fuel to get to the more distant airport, the following four answers are possible:

1) You are justified in believing F, but no longer justified in believing that your (1^{st}-order) evidence supports F.

2) You are justified in believing F and justified in believing that your (1^{st}-order) evidence supports F.

3) You are not justified in believing F, and not justified in believing that your (1^{st}-order) evidence supports F.

4) You are not justified in believing F, but you are justified in believing that your (1^{st}-order) evidence supports F.

4) doesn't seem plausible; even if you can't actually bring yourself to believe F, being justified in believing your evidence supports F prima facie justifies you in believing F.

However none of the other answers seem to be entirely adequate either. It may seem, as in 1, that you could still be justified in believing F – in case your calculation was actually right – but no longer have strong enough reason to believe that the calculation was right. However, this would also mean that you could justifiably believe "F, but my overall evidence doesn't support F." Feldman (2005, 110-111) argues that it is impossible for this belief to be both true and reasonable since the second conjunct undermines the reasonableness of the first conjunct. (Cf. Bergmann 2005, 243; Gibbons 2006, 32; Adler 2002.) And if you were aware of having this belief then you would be believing something that you know is unreasonable if true. You would be, in the view of Feldman and others, disrespecting the evidence. The state in which you believe "F and my evidence does not support F" is a case of "level-splitting", also called *epistemic akrasia*, because you believe you ought not to have a particular belief state but you have it anyway.

The second reply – you are justified in believing F and justified in believing that your evidence supports F – might seem reasonable in some cases, for example if the evidence about one's evidence comes in the form of skeptical philosophical arguments, which one may think are too recherché to command revisions in our everyday beliefs. But this attitude hardly seems acceptable in general since it would mean never giving ground on a first-order belief when presented with evidence that you may be wrong about what your evidence implies. When flying airplanes this kind of rigidity could even be hazardous. However, Feldman counts the second reply as a possible way of respecting the

evidence; it might be fitting not only when faced with radical skeptical arguments but also in cases where one's initial view of what the first-order evidence supports is actually correct.

The third reply, that after noting the altimeter evidence one isn't justified in believing that one's evidence supports F and also isn't justified in believing F has the virtue of caution but also the consequence that the altimeter evidence deprives you of justification for believing F even if you do not suffer from hypoxia, which Feldman takes to be problematic. However, this response, unlike the first answer, respects the higher-order evidence; the altimeter evidence gives you some reason to believe that you might suffer from hypoxia, which gives you some reason to believe your evidence does not support F. The misfortune of being deprived of your knowledge even if you don't actually have hypoxia is an instance of the familiar misfortune of misleading evidence in general. However, as we will see shortly, misleading higher-order self-doubting evidence is distinct from other higher-order evidence, and some recent authors have been led by this to the view that option 1 above – akrasia – can be more rational than option 3 in some cases.

Notably, in both of the replies that Feldman counts as possible ways of respecting the evidence, 2) and 3), the first-order and higher-order attitudes match; one either is justified in believing F and justified in believing one's evidence supports F, or isn't justified in believing F and also isn't justified in believing one's evidence supports F. On getting evidence suggesting one's evidence does not support one's conclusion, one should either maintain that it does so support and maintain the first-order belief – be "steadfast" – or grant that it might not support one's first-order belief and give up the latter – be "conciliatory". If one thinks that which of these attitudes is the right response varies with the case, then the "total evidence" view will be attractive. On this view whether the first order should concede to the second depends on the relative strength of the evidence at each level. (Kelly 2010)

In the conciliatory cases, self-doubting higher-order evidence acts as a defeater of justification for belief, which raises the question of its similarities to and differences from other defeaters. In John Pollock's (1989) terminology, some defeaters of justification for a conclusion are rebutters, that is, are simply evidence against the conclusion, while other defeaters are undercutters; they undermine the relation between the evidence and the conclusion. (These are also referred to as Type I and Type II defeaters.) The pilot we imagined would be getting a rebutting defeater of her justification for believing that she had enough fuel for an extra 50 miles if she looked out her window and saw fuel leaking out of her tank. However, if the altimeter reading is a defeater, then as evidence about whether she drew the right conclusion from her evidence it is definitely of the undercutting type.

All undercutters are evidence that has implications about the relation between evidence and conclusion, and to that extent are higher-order evidence. But the higher-order evidence that leads to self-doubt is distinct from other undercutting-type evidence. In the classic Type II defeater case one's justification for believing a cloth is red is that it looks red and then one then learns that the cloth is illuminated with red light. This evidence undermines your justification for believing that the cloth's looking red is sufficient evidence that it is red, by giving information about a feature of the lighting that gives an alternative explanation of the cloth's looking red. This is higher-order evidence because it is evidence about the cause of your evidence, and thereby evidence about the support relation between it and the conclusion, but higher-order evidence in the cases of the doctor and the pilot are not about how the evidence was caused, and not directly about how matters in the world relevant to one's conclusion are related to each other.

Self-doubting defeaters are about agents and they are in addition agent-specific. (Christensen 2010a, 202) They are based on information about you, the person who came to the conclusion about that support relation, and have direct negative implications only for your conclusion. In the case of the cloth, anyone with the same evidence would have their justification undercut by the evidence of the red light. The evidence that the doctor is underslept, however, would not affect the justification

possessed by some other doctor who had reasoned from the same evidence to the same conclusion using the same background knowledge. The evidence that the pilot is at risk of hypoxia would not be a reason for a person on the ground, who had reasoned from the same instrument readings to the same conclusion, to give up the belief that the plane had enough fuel for fifty more miles.

Christensen argues that the agent-specificity of self-doubting higher-order evidence requires the subject to "bracket" her first-order evidence in a way that other defeating evidence does not. He thinks this means that in no longer using the evidence to draw the conclusion, she will not be able to give her evidence its due. (Christensen 2010a, 194-196) In contrast, in the case of the red light and other cases not involving self-doubt, once the redness of the light is added to the evidence, discounting the appearance of the cloth doesn't count as failing to respect that evidence because one is justified in believing it is no longer due respect as evidence of redness. However, arguably, the difference is not that the self-doubter must fail to give the evidence its due. In the higher-order self-doubt cases we have seen the undercutting evidence does not give the subject reason to believe the evidential relation she supposed was there is not there. It gives reason to think that she doesn't know whether the evidential relation is there, even if it is. If it isn't then in bracketing her first-order evidence she isn't failing to give it its due; it isn't due any respect. Because self-doubting defeating evidence concerns the subject's knowledge of the evidential support relation and not the relation itself it appears weaker than typical defeating evidence. However it is potentially more corrosive because it doesn't give the means to settle whether the evidential relation she endorsed is there and so whether the first-order evidence deserves respect.

If the pilot gives up the belief in F and she had been right about the evidential relation, then she will have been the victim of a misleading defeater. Misleading defeaters present well-known difficulties for a theory of justification based on the idea of defeat because Type II defeaters may be subject to further defeaters indefinitely. For example, if one had learned that the light illuminating the cloth was red via testimony, the defeat of one's justification for believing the cloth was red would be defeated by good evidence that one's source was a pathological liar. If we say that justified belief requires that there be no defeaters then that leads us to disqualify cases where a misleading defeater exists, and subjects will lose justification they might have had, even if the misleading defeaters that are distant facts the subject isn't aware of. But if we refine the view to say that only defeaters for which no defeater exists will undermine justification then a subject will count as justified even if she ignores evidence that looks like a defeater for all she knows, because of the existence of a defeater defeater she doesn't know about. **Undefeated, only defeated defeaters, for all she knows.** In general we will face the question how many of the existing defeater defeaters matter to whether we have a justified belief. (Harman 1973, Lycan 1977)

If despite being at an altitude of 10,500 feet our pilot did do the calculation correctly, then her first-order evidence deserved her belief and her evidence from her altitude and the phenomenon of hypoxia was a misleading defeater. It was good reason to worry that her blood oxygen was low, but it might not have been low, and it would be possible in principle for her to get further evidence that would support this view, such as from the reading of a finger pulse oximeter. Misleading defeaters are not new, but few would be tempted to say in the case where one gets evidence that the light is red that it would be rational for the subject to believe both "my evidence doesn't support the claim that the cloth is red" and "the cloth is red". However, for self-doubting type II defeaters several authors have claimed that such level-splitting can be rational.

For example, Williamson (2011) has argued that it is possible for the evidential probability of a proposition to be quite high, while it is also highly probable *that* the evidential probability is low. For instance one's evidence about oneself might indicate that one has made a mistake evaluating one's evidence, a kind of mistake that would lead one to believe an unsupported conclusion, F. One evaluates the evidential probability of F as high because of one's view of its evidence, but thinks F might well be true without one's belief in it being knowledge.

Another way of arguing that it can be rationally required to respond to a real support relation – that is, for the pilot, to believe F – even when one has evidence it might not exist, and so, should also believe one's evidence does not (or might not) support that belief, is with the thought that a rational norm does not cease to apply just because a subject has evidence she hasn't followed it. (Weatherson 2008, ms.; Coates 2012) This rationale would not sanction akrasia for one who learned the light was red, because the point is restricted to cases where the defeating evidence concerns the subject; we saw above that that makes the defeating evidence weaker, and it is weaker in just the right way to support this approach.

Another way of arguing that akrasia can be rational is to take the existence of a support relation as sufficient for justification of belief in a proposition whether the subject has correct beliefs about that support relation or not (Wedgewood 2011). This is motivated by externalism about justification (crossref SEP: Internalist and Externalist Conceptions of Epistemic Justification), which might be more plausible for justifications subject to self-doubting higher order evidence because it is weaker than other undercutting evidence. In a different tack, it has been argued that a general rule that takes negative self-doubting higher-order evidence to always exert some defeating force on first-order beliefs will be very hard to come by. Because the subject is being asked to behave rationally in the face of evidence that she has not behaved rationally, she is subject to norms that give contradictory advice, and fully general rules for adjudicating between such rules are subject to paradoxes (Lasonen-Aarnio 2014).

Possibly the only thing harder than defending a fully general rule requiring the first-order and second-order to match is accepting the intuitive consequences of level-splitting or akrasia. In this situation one believes a certain belief state is (or might be) irrational but persists in it anyway. Horowitz (2014) has defended the Non-Akrasia Constraint (also sometimes referred to as an Enkratic Principle) which forbids being highly confident in both "q" and "my evidence does not support q", in part by arguing that allowing akrasia also delivers highly counter-intuitive follow-on consequences in paradigm cases of higher-order evidence. For example, if our pilot maintains confidence that F, she has enough fuel, how should she explain how she got to a belief in F that she thinks is true when she also thinks her evidence does not support F? It would seem that she can only tell herself that she must have gotten lucky.

She could further tell herself that if the reason she persisted in believing F despite the altimeter reading was that she did in fact have low blood oxygen, then it was really lucky she had that hypoxia! Otherwise in evaluating her total evidence correctly she would have come to a false belief that not-F. Reasoning so, the pilot would be using her confidence in F as a reason to believe the altimeter reading was a misleading defeater, which does not seem to be a good way of finding that out. Moreover, if she did this argument a number of times she could use the track record so formed to bootstrap her way to judging herself reliable after all. (Christensen 2007a,b, White 2009, Horowitz 2014. For general discussion of what is wrong with bootstrapping, see Vogel 2000 and Cohen 2002.) Akrasia also sanctions correspondingly odd betting behavior.

To the extent that New Rational Reflection (of the previous section) calls for a match between first-order and second-order beliefs it counts as a Non-Akrasia principle. However, this particular matching requirement is subject to several problems about evidence that are brought out by Lasonen-Aarnio (2015). It requires substantive assumptions about evidence and updating our beliefs that are not obvious, and does not appear to respect the internalism about rationality that apparently motivates it, namely that one's opinions about what states it is rational to be in match what states one is actually in. Moreover it does not seem that New Rational Reflection could embody the attractive idea that in general a subject may rationally always be uncertain whether she is rational, i.e., even the ideal agent can doubt that she is the ideal agent – an idea that RatRef failed to conform with and that led to the formulation of this new principle. This is because New Rational Reflection must assume that some things, such as conditionalization, cannot be doubted to be

rational, i.e., to be what the ideal agent would do. It is not clear that we should have expected that everything can be doubted at once (Vickers 2000, Roush et al. 2012), but this is an ongoing area of research (Sliwa and Horowitz 2015).

Another problem some have seen with any version of Rational Reflection is that it ultimately doesn't allow the subject to remain unsure what degree of belief it is rational for her to have. It forces her to match her first-order degree of belief to a specific value, namely, a weighted average of the degrees of belief that she thinks it might be rational to have. It collapses her uncertainty about what is rational to a certainty about the average of the possibilities and forces her to embrace that precise value. This doesn't allow a kind of mismatch or akrasia that might be the right way to respond to some higher-order evidence, where one is confident that q and also thinks it likely that the evidence supports a lower confidence than one has but is unsure what that lower confidence should be. Maybe one should not defer to the higher-order evidence in this case, because one is not sure what its verdict is. (Sliwa and Horowitz 2015) See the higher-order calibration approach below for a way of representing this uncertainty that can give a justification for thinking that matching to averages is rational.

The evidential approach locates the authority that second-order information about our judgment has over us in the idea that it is evidence and we should respect our evidence. A state of self-doubt on this view is confidence both that q and that one's evidence may well not support q. This state does not make the subject inconsistent, but is a state of level-splitting or akrasia. Matching on this view is constituted by agreement between one's level of confidence in q and how far one thinks one's evidence supports q, and doesn't allow akrasia, but taking an evidential approach does not by itself settle whether rationality requires matching, or under what circumstances first- or second-order evidence should determine one's first order confidence. General rules about how to adjudicate in self-doubting cases between the claims of the two orders of evidence may be hard to get due to paradoxes, and to the need in every instance to hold some features of rationality as undoubtable in order to initiate and resolve one's doubt.

4. Calibration and Higher-order objective probability

Another approach to self-doubt explains the authority that second-order evidence sometimes has over one's first-order beliefs by means of the idea that such evidence provides information about the relation of one's first-order beliefs to the way the world is which one is obligated to take into account. That is, evidence like the altitude reading, sleep-deprivation, and empirical studies of the unreliability of eyewitness testimony, provides information about whether your beliefs are reliable indicators of the truth. We take the reading of a thermometer no more seriously than we regard the instrument as reliable. Our beliefs can be viewed as readings of the world and treated the same way. (Roush 2009, White 2009, Sliwa and Horowitz 2015)

a. Guess Calibration

One way of formulating a constraint that says we should be no more confident than we are reliable is by requiring *guess calibration* (GC):

If I draw the conclusion that q on the basis of evidence e, my credence in q should equal my prior expected reliability with respect to q (White 2009, Sliwa and Horowitz 2015)

Your expected reliability with respect to q is understood as the probability – chance or propensity – that your guess of q is true. You may not be sure what your reliability is so you will use an expected value, a weighted average of the values you think are possible, and this should be a prior probability, evaluated independently of your current belief that p.

Self-doubt on this picture would be a state in which you've drawn conclusion q, say because your confidence in q exceeded a given threshold, but also have reason to believe your reliability with respect to q is not as high as that confidence, and such a state would be a violation of GC. Whether this self-doubting state can be coherent when a subject knows her own beliefs depends very much on how the reliability aspect is formulated. If chances and propensities that your guesses are true are determined by frequencies of ordered pairs of guesses of q and truth or falsity of q then self-doubt will make one incoherent here in the way that representing oneself as an anti-expert did above because coherence and expected reliability will be logically equivalent.[2] In any case GC requires matching between the orders in all cases, and tells us that the match is between your confidence and your reliability.

GC makes sense of the intuition in some cases of self-doubt that the subject should drop her confidence. The pilot on looking at the altimeter reading should cease to be so sure she has enough gas for fifty more miles because it gives her reason to think she is in a state where her calculations will not reliably turn out truths. Similarly, the doctor realizing she is severely underslept acquires reason to think she is in a state where her way of coming to beliefs does not reliably lead to true conclusions. The main dissatisfaction with GC has been that it apparently cedes all authority to second-order evidence. In fact in GC's formulation the confidence you should finally have in q does not depend on how far the evidence e supports q or how far you think e supports q, but only on what you think is your propensity for or frequency of getting it right about q, whether you used evidence or not.

There may be cases where second-order evidence is worrying enough that one's first order conclusion should be mistrusted entirely, even if it was in fact soundly made – maybe the pilot and doctor are such cases since the stakes are high. But as we saw above it doesn't seem right across the board to take first-order evidence to count for nothing when second-order evidence is around. Here we can see that by supposing that two people, Anton and Ana, reason to different conclusions, q and not-q, on the basis of the same evidence, Anton evaluating the evidence correctly, Ana not. Suppose both are given the same undermining evidence, say that people in their conditions only get it right 60% of the time. According to GC, rationality requires both of them to become 60% confident in their conclusions. Anton, who reasoned correctly from the evidence, is no more rational than Ana, and has no right to a higher confidence in his conclusion q than Ana who reasoned badly has for her conclusion not-q. (Sliwa and Horowitz 2015)

b. Evidential Calibration

It seems wrong that second-order evidence should always swamp the first-order verdict completely, so the calibration idea has been re-formulated so as to incorporate dependence on first-order evidence explicitly, in the *evidential calibration* constraint (EC):

When one's evidence favors q over not-q, one's credence in q should equal the [prior] expected reliability of one's educated guess that q. (Sliwa and Horowitz 2015)

Your educated guess corresponds to the answer that you have the highest credence in. Reliability of such a guess is defined as the probability that you would assign the highest credence to the true answer if you had to choose, and as above this probability is understood as your propensity to guess correctly. What is used in EC, as in GC, is an expected rather than actual reliability so it is weighted by how likely you think each possible reliability level is. The difference between GC and EC is that in

[2] Probabilistic coherence is logically equivalent to potential calibration in the technical sense contained in the Brier score. (van Fraassen 1983). A sequential forecaster must not revise on the basis of a finite track record that suggests miscalibration, no matter how strongly that frequency suggests it, on pain of incoherence. (Dawid 1982) The higher-order probability approach below avoids this consequence by representing the judgments about the track record explicitly at the second order with a different function for reliability.

the latter the calibration requirement depends explicitly on which conclusion the first-order evidence actually supports. On this principle, Anton, who reasoned correctly with the first-order evidence, is rational to be .6 confident in q rather than not-q because q is the conclusion the first-order evidence actually supports. The contribution of the second-order evidence is to reduce his confidence in that conclusion from a high value to .6.

According to Sliwa and Horowitz EC implies that Ana is not rational to have a .6 confidence in not-q because not-q is not the conclusion the evidence actually favors. It would be rational for her to have .6 confidence that q, the same as Anton. This claim highlights ambiguities in the phrase "one's educated guess that q". Expected reliability is the probability that you would assign the highest credence to the true answer, and the undermining evidence both Anton and Ana were given said that in the conditions they were in they had a 60% chance of their guess being the right answer.[3] If so, then neither of them have enough information to know the expected reliability of a guess *that q*. If the probabilities it appeared they were given, for one's guess about q whatever it is, are to be usable, then the phrase in EC would need to be interpreted "one's educated guess that q or not-q".

The fact that Ana did not actually guess that q makes for a difficulty on either interpretation of the higher-order evidence and EC. Suppose the 60%-probability evidence they were given was indeed only about guesses that q, and that "educated guess that q" in EC refers only to guesses that q. If the word "one's" in the phrase "one's educated guess that q" refers narrowly to the individual EC is being applied to, then EC does not imply anything for Ana, since she did not guess that q. If "one's" refers broadly to anyone who guesses that q in the conditions Anton and Ana were in, then it does follow that what is rational for Ana is to believe q with 60% confidence. However, whether they are given general evidence about guesses that q or not-q, or separate statistics on the success of q-guesses and of not-q guesses, that higher-order evidence would have given Ana no means to correct herself. Because she got it wrong in the first step by incorrectly concluding the first-order evidence supported not-q she will also lack the means to correct herself, that is, to know whether she should be 60% confident in q or in not-q. What EC says it is rational for her to do in the situation is not something she has the ability to do. It might be possible to avoid these difficulties in a reformulation, but they are consequences of the move in EC to add deference to the actual first-order evidential support relation.

EC rules out many cases of bootstrapping that level-splitting views allow. For example, a bootstrapping doctor with evidence that he is unreliable assembles a sterling track record of success in his first-order decisions by judging the correctness of his conclusions by his confidence in those conclusions. He thinks the evidence of his unreliability that he started out with has now been outweighed, so he concludes that he is reliable after all. EC doesn't allow this to be rational because it doesn't allow him to assemble the track record in the first place, since he is obligated in every instance to take into account the expected (un)reliability that he has evidence for. It is unclear, however, that EC similarly rules out bootstrapping for a subject who begins with no evidence at all about her reliability.

The EC reformulation of GC takes a different view of whether rationality requires us to get it right about the first-order support relation or merely to get it right by our own lights, but this question is of course not specific to the topic of the relation of first-order and second-order evidence. For example, in a probabilistic account evidential support relations are dictated completely by conditional probabilities. In a subjective Bayesian version of this picture rationality requires one to have the confidence dictated by the subjective conditional probabilities that follow from one's

[3] We can't assume that one's expected reliability is the same whatever one's confidence is greater in q or not-q, and so, whatever one's guess. Those reliabilities are mathematically independent and human beings' calibration curves (discussed below) tend to vary systematically with level of confidence and proposition, a feature that does not by itself make a subject incoherent.

confidences in other propositions. In an objective version rationality would obligate one to have a confidence that is in line with the objective conditional probabilities. There are other ways of making out subjective vs. objective views of the relevant evidential support relations, and whether we should favor one or the other depends on more general considerations that could provide independent reason to favor one or the other view in the current debate about order relations.

Though this distinction is not specific to the current context it appears to have played a role in some authors' intuitions about level-splitting above. For example, when Weatherson and Coates say that the subject ought to believe what the first-order evidence actually supports because a norm does not cease to apply just because one has evidence one hasn't followed it, they assume that the relevant norm and evidential support relation are objective. Wedgewood's appeal to externalism about justification also takes its bearings from what the first-order evidence actually supports rather than what to one's own point of view it seems to support. A challenge for these approaches that achieve some additional authority for first order evidence over the second order by requiring deference to the actual evidential relation at the first order is to explain why this is an obligation at the first order but a subject need only take into account the *expected* reliability at the second order.

c. Calibration in Higher-Order Probability

Another approach that sees rationality constraints between the two orders as based in taking evidence of one's expected reliability into account derives the constraints top-down from general, widely held, subjective Bayesian assumptions about evidential support, and explicit representation of second-order reliability claims in higher-order objective probability. (Roush 2009) Like approach 2 above it uses subjective conditional probability to express the match that is required between the two orders, but it avoids the consequence we saw in most of those approaches – and in the categorical approach and the other calibration approaches just discussed – that a state of self-doubt combined with knowledge of one's beliefs is incoherent. Unlike the first two calibration approaches it gives an explanation why calibration is part of rationality; it does this by deriving the constraint from another widely accepted assumption, the Principal Principle.

We can write a description of the relation of the subject's belief in q to the way the world is – her reliability – as an objective conditional probability:

$PR(q/P(q)=x) = y$ Calibration Curve

The objective probability of q given that the subject believes q to degree x is y. This is a curve, a function that allows reliability y to vary with the independent variable of confidence, x, with different variables used in order to allow for the possibility that the subject's degree of belief tends not to match the objective probability, and that the level and direction of mismatch can vary with the level of confidence. The curve is specific to proposition q and to the subject whose probability function is P. A subject is calibrated on q, on this definition, if his calibration curve is the line $x = y$.[4]

Calibration curves are widely studied by empirical psychologists who find that human beings' reliability tends on average to vary systematically and uniformly with confidence, with for example

[4] The calibration curve is distinct from the calibration *score*, which is the calibration-related idea most often discussed in philosophy and which has been criticized and largely dismissed, on grounds laid out in Seidenfeld 1985. The calibration score is a root mean squares measure of how far a person's calibration curve is from the $x = y$ line. A single score has a one-many relationship with an infinite number of calibration curves, so a lower calibration score does not uniquely determine how a reduction in the score has been achieved. Consequently one may improve one's calibration score simply by making less informative claims, e.g., having every day the same confidence in rain, a confidence that matches the annual frequency of rain. The use of the calibration curve here is very different. Due to the fact that the subject will be conditionalizing on it, what the person should do with her confidence is uniquely determined, and the improvement is not judged by its effect on her overall calibration score.

high confidence tending to overconfidence, as in eyewitness testimony. Despite the averages found when subjects take tests in controlled settings, the curves also vary with sub-group, individual traits, professional skills, and particular circumstances. All manner of higher-order evidence about a subject's belief-forming processes, methods, circumstances, track-record, and competences are relevant to estimating this function. In real life no one could get enough evidence in one lifetime to warrant certainty about an individual's calibration curve for q in a set of circumstances, but if one is a Bayesian one can form a confidence about what a person's calibration curve is, or what value it has for some argument x, that is proportional to the strength of one's evidence about this, and one can have such a confidence about one's own calibration curve.

On this approach epistemic self-doubt is a state where one is confident and more or less correct that one believes q to degree x, that is, $P(q) = x$, but also has an uncomfortably high level of confidence, say $\geq .5$, that one is unreliable about q at that confidence. That is, one has confidence $\geq .5$ that the objective probability of q when one has x-level of confidence in q is different from x, which we would write $P(PR(q/P(q)=x) \neq x) \geq .5$. Let us say that the different value is y, so $P(PR(q/P(q)=x) = y) \geq .5$, $y \neq x$. Whether or not the reason for this unreliability is that one tends to mistake evidential support relations and whether one does or does not think a given evidential support relation obtains, make no general difference to this evaluation which is simply about whether one tends to get things right when doing the sort of thing one did in coming to be confident in q to level x;[5] it is about the relation between one's confidence and the way things are.

On this view, a state of self-doubt involves a combination of states of the following sort:

$P(q) = x$

$P(P(q) = x) = .99$ (high)

$P(PR(q/P(q) = x) = y) \geq .5$, $y \neq x$

You actually believe q to degree x, you are confident (say at .99) that you so believe, and you have an uncomfortably high level of confidence that you are not calibrated for q at x, that the objective probability of q when you are x confident of q is y. This state escapes incoherence for two reasons. One is that one's confidence either about one's degree of belief or one's reliability is not 1, and unlike some conditional probability formulations of self-doubt above, the slightest uncertainty is enough to make it coherent to attribute a large discrepancy between your believed confidence and your believed reliability.

This is made possible by the second factor, that (un)reliability is expressed here as an objective conditional probability, and coherence alone does not dictate how subjective and objective probabilities must relate. This is analogous to the reason that the approach via the maximally rational subject above was able to represent a state of self-doubt as coherent, namely, that in evaluating my own P I compare it to a different probability function. However, in this case the second function is not an expert function that declares unconditionally what value the maximally rational subject's value for q would be, but a calibration function, a conditional probability that tells one what objective probability is indicated by one's subjective probability. One difference between the two approaches is that there are obvious ways to investigate calibration curves empirically, whereas it would be hard to recruit enough maximally rational subjects for a statistically significant study so we tend to be left appealing to intuitions about what seems rational.

[5] What sort of mistake might have been made is relevant to the estimation of the calibration curve for a particular person and proposition and occasion, since the reliability of one's answer will also vary with method. But the particulars of why you are uncalibrated do not make a difference to how you correct for it or with what justification.

Once the defeating information about the relation of a subject's credences to the world is expressed in objective probability it can be represented explicitly as a consideration the subject takes on board in assessing the quality of the degree of belief she takes herself to have in q and resolving the question what her degree of belief should be, thus:

P(q/P(q)=x & PR(q/P(q)=x)=y) = ?

This asks for the degree of belief the subject should have in q on condition that she actually has degree of belief x in q and the objective probability of q given that she has degree of belief x in q is y. This expression is the left-hand side of Self-Respect/Synchronic Reflection with a further conjunct added to its condition. SR doesn't specify what to do when there is another conjunct and so is not suited to explicitly represent the question of self-doubt, which means that the self-doubting examples above are not counterexamples to it. (Roush 2009) However, some in the past have endorsed variants on an unrestricted version of SR (Koons 1992, Gaifman 1986) where the value of this expression is x regardless of what other conjunct might be present:

P(q/P(q)=x & r) = x, for r any proposition Unrestricted Self-Respect (USR)[6]

Dutch book arguments that might give support to SR do not do the same for USR, leaving us with a need to find other ways of evaluating it when r is the statement of a calibration curve.

It is not incoherent but it is baldly counterintuitive to suppose that the subject should have degree of belief x when she believes that her so believing is an indicator that the objective probability of q is not x, and a principled argument can also be made to this effect. (Roush 2009) Unpacking the condition P(q)=x & PR(q/P(q)=x) = y, it seems to say that my credence is x and when my credence is x the objective probability is y, inviting us to discharge and infer that the objective probability is y. If so,[7] then the expression would reduce to:

P(q/PR(q)=y) = ?

which is the left-hand side of a generalization of the Principal Principle (crossref SEP: David Lewis)

P(q/Ch(q)=y) = y Principal Principle (PP)[8]

from chance to any type of objective probability. PP says that your credences in propositions should conform to what you take to be their chances of being true, and, admissibility issues

---

[6] Though not noted by authors who have endorsed it, it is evident that USR needs admissibility conditions to rule out obvious counterexamples such as taking r to be q itself. Since the calibration approach being discussed does not depend on USR this can be ignored in this context.

[7] P(q)=x and PR(q/P(q)=x)=y do not by themselves imply PR(q)=y. To discharge that conditional probability requires that PR(P(q)=x) is high, so one must ask under what conditions the credence being x makes the objective probability high that the credence is x. Under what conditions does A imply that A is objectively probable? This is an interesting question but it can be avoided in the derivation by appealing to a variation on the Principal Principle (Vranas 2004),

P(q/B & Ch(q/B)=y) = y Conditional Principle (CP)

to which no one has objected. The generalization of the Conditional Principle from chance to any objective type of probability is:

P(q/B & PR(q/B)=y) = y General Conditional Principle (GCP)

This says that the credence in q given that B is true and that the objective probability of q given that B is true is y, is y. Here too there are questions of admissibility, but as with PP it's easy to expect that there exists a useful domain in which CP and GCP are true.

Cal below is an instance of GCP:

P(q/(P(q)=x & PR(q/P(q)=x)=y)) = y Cal

[8] This formulation of PP suppresses a conjunct r of the kind that USR above has, meaning that it also requires admissibility conditions to identify the domain in which it is true. Those would be carried over in the use of PP in what follows here.

notwithstanding, it is hard to deny that there exists a domain in which the Principal Principle is compelling, and surely one where the generalization to any type of objective probability is too. If so then the answer to the question what the subject's credence in q ought to be in light of her consideration of information about her reliability is:

$P(q/(P(q)=x \ \& \ PR(q/P(q)=x)=y)) = y$ Cal

Cal says that your credence in q given that your credence in q is x and the objective probability of q given that your credence in q is x is y, should be y.

Cal is a synchronic constraint, but if we revise our credences by conditionalization then it implies a diachronic constraint:

$P_{n+1}(q) = P_n(q/(P_n(q)=x \ \& \ PR(q/P_n(q)=x)=y)) = y$ Re-Cal

This calibration approach tells the subject how to respond to information about her cognitive impairment in every case. It uses the information about herself to correct her belief about the world. Intuitively it is a graded generalization of the thought that if you knew of someone (or yourself) that he invariably had false beliefs, then you could gain a true belief by negating everything he said.

Cal and Re-Cal give an explicit characterization of self-doubt and justification of a unique and determinate response to it on the basis of deeper principles that are compelling independently of the current context. Cal follows from only two assumptions, first that probabilistic coherence is a requirement of rationality, and second that rationality requires one's credences to align with what according to one's evidence are the objective probabilities. Re-Cal comes from further assuming that updating our beliefs should occur by conditionalization.

Although self-doubt under the current definition of it is not an incoherent state, Cal implies that rationality always requires a resolution of the doubt that brings matching between the two levels, and tells us that the matching consists in the alignment of subjective and perceived objective probabilities. High confidences in "q", "I have confidence x in q", and "the objective probability of q when I have confidence x in q is low" are not incoherent, but they do violate the Principal Principle. Re-Cal tells us how to get back in line with PP.

Though Re-Cal has us conditioning on second-order evidence, the adjustment it recommends depends on both first- and second-order evidence and does not always favor one level or the other. How much authority the second-order claim about the reliability/calibration curve has depends very much on the quality of one's evidence about it. This can be seen by imagining being uncertain about, e.g., one's calibration curve, i.e., $P(PR(q/P(q) = x) = y) < 1$, and doing a Jeffrey conditionalization version of Re-Cal. (Roush ms.) But even in case one has perfect knowledge of one's calibration curve, the role of the first-order evidence in determining one's first-order belief is ineliminable. The verdict, the level of confidence, that the first order gave you for q is the index for determining which point on the calibration curve is relevant to potentially correcting your degree of belief. To understand why this is far from trivial, recall that the curve can in principle and does often in fact have different magnitudes and directions of distortion at different confidences. The verdict's dependence on the first-order evidential support relation is different from that of EC in another way, since it uses not the objective support relation at the first order but the consequences of the subject's take on it. Thus, Ana above would not be left not knowing how to make herself rational.

The fact that the update proceeds by conditionalization means that all of the kinds of evaluation of evidence that conditionalization imposes come along with that. Misleading self-doubting defeaters troubled some authors above and led them to level-splitting views, but they are handled by Re-Cal as conditionalization always handles them. Self-doubting defeaters are processed at face value as relevant to the calibration curve in proportion to their quality as evidence. Convergence theorems tell us that if the world isn't systematically deceptive then the misleading defeaters will be washed

out, that is, defeated by some other evidence, in the long run. In some cases that will happen only long after we're all dead, but if one views that as inadequate then that is a dissatisfaction with subjective Bayesianism, and is not specific to its usage here.

The approach to epistemic self-doubt in terms of higher-order probability allows the state of self-doubt to be rational (coherent), and to be rationally resolved. Cal expresses a requirement of matching between the two orders in all cases, though it does not imply that *attributing* a mismatch to oneself is incoherent. Neither order is always dominant; both orders always make a contribution to determining the resolution at the first order of conflicts between orders, and their relative contribution depends on the quality of the evidence at each order. Cal and Re-Cal explain why one should revise in light of higher-order evidence, when one should, by reference only to probabilistic coherence, the Principal Principle, and conditionalization. Cal and Re-Cal are general and make available all of the resources of the Bayesian framework for analysis of higher-order evidence. A further notable fact about the framework is that Re-Cal allows for cases where news about one's reliability should increase one's confidence, which would be appropriate for example in cases, easy to imagine, where one acquired evidence that one was systematically *under*confident. Thus, it is possible for second-order evidence to make it rational to be not only steadfast or conciliatory, but even emboldened.