Second Guessing: A Self-Help Manual[1]

Sherrilyn Roush

Abstract

I develop a general framework with a rationality constraint that shows how coherently to represent and deal with second-order information about one's own judgmental reliability. It is a rejection of and generalization away from the typical Bayesian requirements of unconditional judgmental self-respect and perfect knowledge of one's own beliefs, and is defended by appeal to the Principal Principle. This yields consequences about maintaining unity of the self, about symmetries and asymmetries between the first- and third-person, and a principled way of knowing when to stop second-guessing oneself. Peer disagreement is treated as a special case where one doubts oneself because of news that an intellectual equal disagrees. This framework, and variants of it, imply that the belief as typically stated that an equally reliably peer disagrees is incoherent, and so pure rationality constraints without further substantive information cannot give an answer as to what to do. The framework also shows that treating both ourselves and others as thermometers in the disagreement situation does not imply the Equal Weight view.

 "Second-guessing" is a pejorative term for something we think you should stop doing. We

know from ordinary life that those we call "second-guessers" tend to fall into paralyzing

regresses. The psychological literature on the various maladaptations and perils of chronic

self-doubt about one's judgments is clear. For example:

> High self-doubters, because they give diminished weight to their own
> interpretations and perspectives, are, so to speak, not well 'centered.' We believe
> that this is why they are prone to having their moods buffeted about by changes in
> their immediate circumstances. (Mirels, L., quoted in Dittman 2003. See also
> Mirels et al. 2002.)

High judgmental self-doubters tend to suffer not only from mood swings, but indecisiveness, proneness to change opinion with exposure to the opinions of others, procrastination, low self-esteem, and anxiety, among other symptoms. One intriguing consolation is that they are not found to differ from others on measures of intelligence, which suggests they are getting something right.

The list of the chronic self-doubter's sorry symptoms makes it seem as if ridding ourselves of sensitivity to our cognitive imperfection could be a mental health panacea. However, we also know that awareness of whether or not we are fit to judge can be a very helpful thing. I will argue here that a self-doubting responsiveness that is pathological when taken to an extreme is, properly applied, a requirement of rationality. My explicit formulation of how we are obligated to take our fallibility into account will explain the benefit of doing so and the mistake the pathological self-doubter is making. My account also has implications for current discussions of epistemic self-respect, the symmetries and asymmetries between first- and third-person judgments, and peer disagreement. Peer disagreement is the special case of judgmental self-doubt where what prompts the consideration that one might be wrong is discovery that a person one regards as an intellectual equal disagrees.

The issue of how properly to be responsive to our cognitive imperfection is significant in a number of ways. One is a difficulty understanding how to admit the fallibility of science without falling into a skepticism that regards all theories as equally plausible. This difficulty is exploited, for example, in public discourse when creationists infer from the fallibility of science – which we cannot but admit – to the claim that their views have equal rights with evolutionary biology to be taught in public school science classes. Of course, the fact that we might be wrong does not mean that every view is equally likely, but it is not trivial to say exactly why. The resemblance of the pessimistic induction over the history of science to the creationist argument is also alarming because that induction is intuitively harder to dismiss. Another project I have is to characterize justified belief and say what it is good for, given that I have taken the view that it is not essential to knowledge. Using the account here of how to be responsive to the possibility

one has erred allows me to define justified belief in such a way as to respect the internalist intuition that justified belief requires us to be checking ourselves, without imposing a requirement that we have conscious access to our reasons. It thereby allows a unified account of perceptually and argumentatively justified beliefs; in different contexts, for different propositions, one achieves the same self-monitoring by different means. I use probability to formulate the one rule that I think governs in all these contexts, and as such the core of this work is a generalization of the Bayesian account of rationality away from extreme idealizations about the relation between beliefs and beliefs about one's beliefs.

The Framework

Information suggesting one should reconsider whether q, despite having no new first-order evidence for or against q, can come in many forms. For a realistic one, suppose you are very confident that q, John is the murderer, from having witnessed the crime directly. The spectacle was so traumatic that his face is seared into your memory forevermore. Your degree of belief in q is .95, say. Then you learn of psychologists' substantial empirical evidence that human beings are generally more unreliable at eyewitness testimony than any of us imagined; when we're confident, we tend to be overconfident. Our question is, What are you supposed to do with your confident belief that John is the murderer in light of this evidence about your unreliability in judging such things, and why is that what you are supposed to do?

If it seems to you that you should reduce your initial high confidence that John is the murderer, then you are endorsing a violation of what David Christensen has called the principle of Self-Respect (SR) (Christensen 2007, 322):

$P(q/P(q) = n) = n$ (SR)

This says, literally, that if you are rational, then your degree of belief in q, given that your degree of belief in q is n, is n. As Christensen puts it: you should not disapprove of your

own current credences. It is hard to see how to deny this principle when it is put that way, but Christensen has misgivings about it too, for reasons similar to the one we just saw in the example, that if you learn that confident people are overconfident about what they eyewitness, and you know that you are confident, then it doesn't seem right for you to sit tight; in such cases the second "n" in SR should be less than the first "n".

The Self-Respect principle has a long history of discussion under the name "Miller's Principle." It is also the synchronic case of van Fraassen's Reflection Principle. From both these directions it has been roundly endorsed, a number of people reporting an inability to imagine how it could be wrong. That there are seemingly easy counterexamples to a principle that seems obvious to many people is, as I will explain, due to an ambiguity in its statement. It could be that in mentioning nothing other than "P(q) = n" in the condition of SR we implicitly assume a proviso, that no other statement for which the subject has a degree of belief is together with "P(q) = n" probabilistically relevant to q. In this case the principle surely is hard to imagine wrong, since it is the weak claim that your degree of belief in q being n is not *by itself* a reason to have a different degree of belief in q than n. This I will call "Restricted Self-Respect" (RSR):

$P(q/P(q) = n) = n$ *if no other statement of probability[2] for which* P *has a value is*

relevant (with $P(q) = n$) *to* q   (RSR)

The situation above with the eyewitness testifier does not fulfill the proviso. You are not contemplating reducing your confidence that John is the murderer merely in virtue of noticing that you are confident that he is. Your loss of nerve is coming, if it is, from having in addition learned that confident eyewitness testifiers are overconfident, and recognizing that you are a confident eyewitness testifier. Thus, if you have a different degree of belief that John is the murderer in these circumstances then you are violating only a much stronger principle than the one just described, a principle that says:

---

[2] "Statement of probability" refers to a statement of the form $P(A) = x$ or $PR(A) = x$, so, for example, the statement q is not referred to by this proviso. q is relevant to q, and what value P assigns to it is referred to by "$P(q/P(q) = n)$", but q functions in the background, not being taken to have probability 1, as a term in the condition would be, but taken to have whatever probability value it has in P.

$P(q/P(q) = n \cdot r) = n$ *for* r *any statement of probability for which* P *has a value*[3]   (USR)

I will call this stronger principle "Unrestricted Self-Respect" (USR). In our case this principle would say that your degree of belief that John is the murderer given that your degree of belief that he is is .95, should be .95 *no matter what* your other beliefs. That is, even if you also believe that psychologists firmly established that eyewitness testimony is unreliable. It is in so far as a Self-Respect Principle says "no matter what" instead of "as long as there are no other relevant beliefs" that cases like that with the eyewitness testimony research are counterexamples. Surely, though, cases like this illustrate that our respect for ourselves, just like our respect for others, should not be unconditional. The appropriate rationality requirement here should not be USR (or SR), but a general principle of which RSR is a special case, a general principle that explains when and how we should be sensitive to news of our cognitive weaknesses, and why.

The principle that is intuitively obvious is RSR. The principles that have counterexamples are USR, and Christensen's SR because he understands it as unrestricted over background beliefs, and beliefs about unreliability could occur in either place. Thus I would explain the tensions Christensen has uncovered in considerations for and against his SR as due to treating RSR and SR as if they were the same. The standard counterexamples don't create tension once the principles are distinguished. Christensen (2007) has shown that perfect knowledge of our beliefs implies SR, where this perfect knowledge requires having extreme confidence, for every degree of belief about every proposition, about whether I have that degree of belief or not (Confidence), and requires that I be correct in all those beliefs about my beliefs (Accuracy). Thus, in rejecting SR I

---

[3] This principle should be understood as making the claim whether r is part of the condition or part of the background. It is also understood as applying only on assumption the expression "$P(q/P(q) = n \cdot r)$" is not otherwise undefined, e.g. because of an incoherent condition. Otherwise USR would be trivially false.

also reject Confidence and Accuracy as rationality constraints.[4,5] My retention of RSR is in this respect unproblematic because RSR does not imply Confidence and Accuracy. In addition, we will see below that perfect or necessary knowledge of our own beliefs is not required for the unity or epistemic functioning of the self, properties whose preservation has I think motivated some to make these extreme assumptions about knowledge of our own beliefs.

That rejection of Confidence and Accuracy as rationality requirements allows denying USR and SR and maintaining RSR and unity of the self is not a specific reason to reject these perfect self-knowledge conditions. The mere fact that we do not actually fulfill the conditions is not a reason either, since we are speaking of normative constraints of ideal rationality. My specific reason for rejecting them is that the fact that I have a certain degree of belief, just like the fact that someone else has a certain degree of belief, is an empirical matter. My having perfect knowledge of these things is not a requirement of rationality for the same reason my having perfect knowledge of how many people are infected with swine flu is not a requirement of rationality. In the following discussion I will assume that the subjects in question have knowledge of their beliefs that is as good as you like, not perfect, because of the contingent fact that one wouldn't have the quandaries in question, or they wouldn't be interesting, if one didn't have this knowledge.

As I have explained, USR claims that no statement of probability could be relevant to q when combined with the statement of my degree of belief in q, $P(q) = n$. The non-

---

[4] It is in my view purely an artifact of representing these matters via probability that knowledge of what one's degrees of beliefs are implies that those beliefs are what they should be, an idea that is bizarre at best. The implication follows because of the fact that if the self-knowledge is perfect that must be represented via extreme probabilities, and 0 and 1 values automatically induce independence relations. The fact, displayed in the rest of this paper, that there is a great amount of coherent structure if we leave out the extreme cases is confirmation of my claim of artifact. Another case of extreme probabilities hiding structure by fiat is discussed in *Appendix*.

[5] In fact, it appears that it is not sufficient for me merely to deny the conjunction of Confidence and Accuracy as general principles. The rationality constraints I end up defending in this paper appear to be coherent only for those *cases* of q for which either Confidence or Accuracy fails. There is then a decision point: one might allow that it is rational to be perfectly confident about what one's degree of belief in a proposition is, and see the constraints I am presenting as applying only when one is not perfectly confident, or one might reject its ever being rational to be perfectly confident of what one's beliefs are. I favor the latter option, which would follow from my claim (below) that whether one has a particular degree of belief or not is an empirical matter, taken together with radical probabilism – the view that one should have extreme degrees of belief only in logical truths and falsehoods. However, I don't need to take a stand on this question in this paper.

pathological[6] cases in which this is false are ones like your confident belief above that John is the murderer, when you also learn about the unreliability of judgments such as yours. The key to generalizing RSR is to represent that belief you come to have about your reliability *explicitly*. This will both yield the new rule, and justify rejection of USR.[7] I will represent a claim that a subject has reliability level z when believing q to degree n as follows:

$PR(q/P(q) = n) = z$

to be read "The objective probability of q given that the subject believes q to degree n is z." Since z is a probability, it is a number between 0 and 1 inclusive. Now your quandary when presented with the psychologists' discoveries is represented by your believing a conjunction:

$P(q) = n . PR(q/P(q) = n) = z$

You believe *that* you have degree of belief n in q and *that* the objective probability of q given that you believe it to degree n is z. Our question what believing these things means for your degree of belief in q is: What is the value of the following conditional probability?

$P[q/P(q) = n . PR(q/P(q) = n) = z] = ?$!

That is, what is the right degree of belief to have in q given that you have degree of belief n and the objective probability of q given that you have degree of belief n is z? In the murderer case, what is the right degree of belief that John is the murderer given that you learn your reliability at eyewitness testimony is less than the degree of belief you now have? Notice that this is an instance of USR:

---

[6] Brian Skyrms (1980) already showed that the principle I call USR, and that in that literature is called "Generalized Miller's Principle," has counterexamples, that would also work against SR. However, those he described are pathological – "scarce and peculiar," as he put it – which means that regarding them as inadmissible is a sensible strategy in formulating a rationality constraint that balances simplicity and informativeness. The counterexamples I am dealing with here are commonplace, and with the results of experimental psychology they are growing every day. This is why I claim a generalization needs to be made that takes them into account as admissible cases.

[7] I am using USR in this argument for ease of presentation. The argument can be adapted to justify rejection of RSR because "$PR(q/P(q) = n) = z$" can be represented as in the condition, or in the background with probability 1, indifferently.

P[q/P(q) = n . PR(q/P(q) = n) = z] = **?**

USR implies that the value is n – you should have the degree of belief in q that you believe yourself to have *no matter what*.

This does not seem right intuitively, and a version of the familiar Principal Principle will explain why. Notice that with a natural assumption[8], the conjuncts of the condition imply an objective probability for q, so ! should have the same value as the expression:

P(q/PR(q) = z)

The Principal Principle says:

P(q/PR(q) = z. r) = z,   (PP)

where r is any (admissible) probability statement. This says that your degree of belief in q given that you regard q as having objective probability z, should be z. That is, your subjective degree of belief should conform to what you think the objective probability is. We apparently have no need for inadmissible r in our cases, so the Principal Principle says that the term in question equals z:

P(q/PR(q) = z) = z

implying that:

P[q/P(q) = n . PR(q/P(q) = n) = z] = z

Unrestricted Self-Respect said that the value was n. PP tells us that the value is z. There is no reason to think that n and z are in general the same.

Allowing that it is possible for n to be unequal to z without violation of rationality forces us to choose between the PP and USR. PP is less fishy, and it also explains our intuitions about taking information about your reliability into account, whereas USR conflicts with them. Thus, I advocate rejecting Unrestricted SR, while maintaining PP and

---

[8] P(PR(P(q)=x) = 1/P(q)=x) = 1, which is an instance of P(PR(A) = 1/A) = 1. I.e., you are certain given A that the objective probability of A is 1.

Restricted SR. This implies a general rationality constraint that allows us to see, fully generally, what rationality requires when we are faced with news about our cognitive prowess (on the natural assumption in fn. 8, and assuming that one fails to have either perfect confidence or perfect accuracy about one's degree of belief in q):

$$P[q/(P(q) = n \,.\, PR(q/P(q) = n) = z)] = z \qquad\qquad \text{Cal}$$

The useful upshot of this is in a principle of conditionalization:

$$P_f(q) = P_i[q/(P_i(q) = n \,.\, PR(q/P_i(q) = n) = z)] = z \qquad \text{Re-Cal}$$

When you come to believe both that your degree of belief in q is n and that q is z probable when your degree of belief in q is n, then believe q to degree z. In other words: change your confidence to your believed reliability. The end state of that step of updating is what I call calibration. Psychologists and statisticians formulate what calibration is in a variety of distinct ways, and all, I think, distinct from this one, but the general idea of calibration is that your confidence should match your reliability (which psychologists tend to call "accuracy"). Here, since we are dealing with rationality constraints, the requirement is that your (first-order) confidence in q should match what you *believe* your reliability about q-like matters to be rather than what your reliability level in fact is. Re-Cal is a rule that can be expected to be beneficial, provided you have sound beliefs about your reliability, since it gets you into line with the Principal Principle when you notice things that indicate you are out of line.

The *PR* term in Re-Cal is meant to be the subject's estimate of her reliability on q given every relevant thing she knows and learns about her cognitive faculties. For example, imperfection of reasoning, unreliability of sense perception, skewed gathering of evidence, anything relevant to the way she came to make her judgment about q. The judgment of one's reliability even in a single case involves a lot of factors, e.g., the probability that you have been given a mind-altering drug, the fraction of people who make cognitive mistakes when they have been given that drug, how similar you are to the average person in the effect of drugs on you, whether you used this time a process to which that drug effect is relevant.

Notice a few quantitative relationships involved in this formula. It does not say that news about your reliability will always create an obligation to dial your first-order confidence *down*. If the news is that you are underconfident then Re-Cal tells you to dial up. It may be you are already calibrated; then you do nothing. We also see the difference between creationists and pessimists as a difference in quantity. To admit you are fallible is to admit a small deviation from perfection which does not imply every view is equally likely, whereas the pessimist typically argues that false theories predominate in the history of science. How bad the news of fallibility is for your first-order confidence is related to how far that confidence deviates from what you learn your reliability to be.

Implications: You can believe you're stupid if you try.

One might admit that the condition in Re-Cal makes sense when n does not equal z, as long as it is said of someone else's beliefs. But surely it is somehow incoherent for me to be attributing this to myself. The framework allows for n very high and z very low. Isn't this like attributing to myself anti-expertise about q, which Egan and Elga (2005) have shown is not rational when you have high confidence in q and decent knowledge of what your beliefs are?

It is like, but very different. They have shown that you cannot coherently self-attribute anti-expertise (while maintaining confident belief and decent self-knowledge of your beliefs) on two ways of representing anti-expertise and self-ascription of it, but those representations are not sensitive to the fact that reliability is a property *of* beliefs, and that ascribing it or its negation to myself is a second-order matter. They appeal to intuitions about a subject's beliefs about her beliefs, but do not represent this structure in the formal account. Once one represents all of that explicitly as I have, no problem about self-attribution arises from first-order or within-order coherence constraints or from RSR.

Drawing this out in more detail, in my formulation the situation of being confident *that* you have a certain degree of belief in q and the objective probability of q when you have that degree of belief is a different value, are represented as we saw:

$P[P(q) = n . PR(q/P(q) = n) = z] = 1 - \varepsilon$

The conjunction that you are highly confident about implies that the objective probability of q is z, not n. You are pretty much right, we assume with Egan and Elga, in your high confidence that your degree of belief in q is n, so you are in a situation where:

$P(q) = n \pm \varepsilon$ and $P(PR(q) = z = 1$ STET

That is, you do believe q to roughly degree n and you think the objective probability of q is z.[9] This is not good, of course. It means you are in violation of PP. However, it is not an incoherence because these beliefs do not put the subject in violation of the axioms. It is formally obvious that the "$PR(q) = z$" term might as well be a B for all the axioms are concerned about the relation between your two beliefs here; that term cannot interact with q via the axioms governing P. (See *Appendix*.) The significance of the distinction between violating PP and violating within-order coherence in the current topic is that you cannot (normally) update on an incoherent condition – the result will be undefined. You can update on a set of beliefs that only gives you a violation of PP, and Re-Cal shows you how to do it.

It is crucially important for the current topic that we be able to see how it is rational to update on that set of awkward beliefs we could come to have about ourselves, since otherwise there is no coherent and non-arbitrary way to improve our situation on learning of our unreliability. This is analogous to the problem Socrates pressed on Meno, that if he could not admit that he did not know, then he would never be able to learn. What prevented Meno's admission was a defect of character, roughly speaking. Here, what was in danger of preventing the admission of unreliability was an apparent incoherence of the admitted content. If you cannot coherently admit extreme fallibility, then you cannot learn

---

[9] I leave the "1" for simplicity on the second term; having an extreme value for that term does not matter to my argument.

from news of it, so if you can't believe you're stupid, it's not just a curiosity; it's *really* bad. Thus, it is a reassuring consequence of my framework, that you can believe you're stupid if you try.

Implications: Reverse Golden Rule

Some asymmetries have been noted between how it seems we should handle news about our own beliefs and faculties, and how we should handle news about others'. The fact *that* Alice believes p is not by itself an additional reason for Alice to believe p, but the fact that she believes p may be an additional reason for me to believe it, it seems. But Alice's belief that she believes q and my belief that she believes q apparently have the very same content and are both beliefs about a belief. Thus, according to current notions of higher-order evidence, they are the same kind of evidence. Why should they not be treated the same?

In fact in my framework there is a strong sense in which we should treat others' beliefs the same as we should treat our own. The natural response to my discovering Alice's belief is to take the fact that she has that belief exactly as seriously, as evidence, as I judge Alice to be reliable in her belief. In representing how we should update on news about our own reliability explicitly as I have, I have us doing the same thing to ourselves. Suppose I think Alice is unreliable and discover that she believes q to degree .9, and I have no other beliefs relevant to these matters. The natural way to write the conditional probability behind what I should do about Alice's belief is, taking A to be Alice's belief function and "P" to represent my degrees of belief:

$P[q/(A(q) = .9 . PR(q/A(q) = .9) = .5] = .5$ ‡

$P(q/PR(q) = .5) = .5$ by ‡ and PP

If I believe that Alice has degree of belief .9 in q, and believe that when she is confident of q, q is as likely as not to be true, and I have no other relevant beliefs, then by PP my degree of belief in q should be .5.

The rule above for updating on beliefs about my own reliability and degree of belief is the special case of this where A(q) = P(q), that is, where the belief function in question is my own probability function. Because we can say what I should do about my own self-discoveries by substituting my probability function in where the belief function of another would be in the general case, we can say, in one clear sense, that what I should do with respect to my own belief in q is exactly what I should do with respect to Alice's, or any other person's belief in q.

$P[q/(\mathbf{P(q) = n} \cdot PR(q/\mathbf{P(q) = n}) = z)] = z$

$P[q/PR(q) = z)] = z$ (I kept this one with square even though PP above wasn't, because this one is falling out of one with square brackets.)

That is, I should conjoin the belief that I have the belief with a belief about the reliability of that belief of mine, and re-calibrate.

The re-calibration rule has me evaluating and treating myself as if I were another so we could call this the Reverse Golden Rule. This should not be seen as a radical proposal, since it is evident in our phenomenology that when a person reflects on herself there is a clear sense in which the subject splits into two, the subject reflecting and the subject reflected upon. (In my account, of course, a psychological act of reflection is not required.)

How can we explain the asymmetry originally noted between the way Alice should respond to her having a belief and the way I should respond to her having a belief? Intuition said she should not have a response, but that it is possible I should. The first thing to say is that both intuitive cases are underdescribed, -- there are four possible permutations, not two – and that whether belief revisions need to be made depends not on whose belief is discovered, one's own or somebody else's, but on whether there is also a belief about the reliability of that belief. The four possibilities are that I discover Alice has a belief and I do or do not have a reliability judgment about her belief, and Alice discovers that she has a belief and does or does not have a reliability judgment about her belief. If there is a reliability judgment then a change of the discoverer's belief may be needed. If there is not, then no change is called for. Thus, even if I have a belief that Alice believes q,

if I don't also have a belief about her reliability, then I have been given no reason to change my belief. This is just the same way RSR tells us I should treat my own belief. And if I believe I have a belief or Alice believes she has a belief, it is not automatic that I or she should do nothing. If I also have a belief about my reliability or she has a belief about hers, then I or she should revise to that respective reliability level (unless already there).

There need be no worry that this introduces too much symmetry between myself and others. The subject "reflected" upon here is represented by my first-order beliefs, and the "reflecting" subject is my second-order beliefs. The fact that despite this if I am to be a rational subject I should surely also be a *unity* is expressed in my requiring that all of those beliefs of mine form one probability function, that is, that first- and second-order beliefs fulfill all coherence constraints with respect to each other. The function for my beliefs is "P" throughout, whether that occurs at the first order (once) or second order (twice, nested). There is no counterpart structure in the way I should deal with my beliefs about Alice's beliefs. Her belief function is not my belief function – even if her belief function is so well-behaved as to be a probability function, it is not *my* probability function – so my judgments about her beliefs are never second-order. My ascription of a belief to Alice is to this extent no different from my ascribing to her red hair (and if her having red hair were well-correlated with q being true I should take that as a reason to believe q every bit as much as if it were Alice's *belief* that I was taking to be a reliable indicator of q.) Another consequence of this structural difference is that I have no rational obligation to make my beliefs coherent with hers; my beliefs only must be coherent with what I *believe* about her beliefs and reliability.

So, all judgments that could affect my beliefs must be made by me, whether those are judgments about (the world or) my beliefs and reliability or judgments about the beliefs and reliability of others. It's just that that does not by itself imply that the fact *that* I have a given first-order degree of belief (which is distinct from the content of the belief itself) is privileged in how much respect I should accord it as compared to the fact *that* Alice has a given belief. As thermometers we are not per se prioritized. I have both full identity with and full distance from myself.

Implications: Peer Disagreement

In drawing out what I should do on discovering the belief of another I assumed that I had no beliefs relevant to q except the belief that Alice had a high degree of belief in q and, in some scenarios, that she was so-and-so reliable. The situation of peer disagreement is one where that proviso does not hold, because I discover the belief of another when I also have beliefs about my own beliefs and reliability. In the classic example, suppose I and a friend Alice whose reliability I judge equal to my own have dinner and each of us calculates a number for half of the bill. The problem is that we come up with different numbers, she \$43, I \$45. Should I change my belief to 43? Hang tight? Average the two answers? Does Re-Cal have a recommendation?

Suppose I believe to degree .9 that the answer is \$45 – call this "q" – and Alice believes to degree .9 that the answer is \$43 – call this "p". The assumption that I regard us as equally reliable means that I believe: $PR(q/P(q) = .9) = z$ and $PR(p/P_A(p) = .9) = z$, for some z. Thus the question of peer disagreement is the question what I am supposed to believe about the bill given that I have the following set of beliefs:

$P(q) = .9$. $PR(q/P(q) = .9) = z$. $P_A(p) = .9$. $PR(p/P_A(p) = .9) = z$

We have two sets of claims now, one for me, one for her. These imply a simpler conjunction:

$PR(q) = z$. $PR(p) = z$

p and q are incompatible, so if z is > .5, as is assumed in peer disagreement cases, then this expression in the condition of my conditional probability is a contradiction. **What we say to describe the peer disagreement case reduces to two incompatible objective probability statements.** Thus the answer to the question what I should do with my belief under the conditions of peer disagreement as typically described, which becomes the following via algebra:

$P_f(q) = P_i(q/\ PR(q)\ >.5\ .\ PR(p)\ >.5) = ?$

is: undefined.

When we represent explicitly what is actually being said in describing the peer disagreement case – ascribing equal, high reliability, and different beliefs on the very same issue – we see that since the beliefs imagined are outright contradictory, it is not a condition that can be updated on. Further assumptions or information will have to be added if there is to be a value to this conditional probability or a degree of belief to update to. Such further information could tell me that one or the other of us is unreliable or has made a mistake this time. If it is my friend then I should be something in the direction of steadfast. If it is me then I should lean toward being conciliatory. If I get reason to believe $z = .5$ then I should feel half-sure of each answer. It seems that any of the various positions that have been taken in the literature could be the correct answer in a given case, but only if more information about the case is supplied will there be an answer at all.

Of particular interest here is the consequence that treating Alice and myself with equal respect as thermometers, as we are doing in my framework, does not imply the so-called Equal Weight view, according to which my response to Alice should be to adopt a belief in the average of our answers or to come to have equal credence of .5 in both of our answers. In addition, the idea that the fact that even after learning about Alice's belief, I still have my first-order evidence for my original answer may tip the balance in my favor is rejected here; to take my first-order evidence for q into account after coming to my belief and discovering Alice's belief, and take it as giving me an advantage would be double-counting the first-order evidence. The significance of the first-order evidence was already recorded in my degree of belief about the bill, which I came to by conditionalizing on that evidence before learning of Alice's belief.

The fact that there is no answer to the peer disagreement question without more substantive information means that there is not a solution of the peer disagreement case that is dictated by rationality understood as coherence, plus conditionalization, plus Restricted Self-Respect, plus the Principal Principle. There are other ways of representing

peer disagreement using this framework and also a variant definition of reliability, but, as I show in other work, though those formulations lead down different paths the consequence is substantially the same, that further information is needed; rationality alone will not yield an answer.

How do I stop this thing?

I started with the question how it could be exactly that it is good to be responsive to the reliability level of your cognitive faculties, and yet bad, even pathological, to take this too far. What is the difference? My answer is that much in the pathological case can be explained by imagining a few mistakes in executing the very same rule that governs healthy self-doubt, the Re-Cal conditionalization principle. The main way, it seems, that second-guessing oneself gets out of hand is its leading to a regress in which at each step we doubt our judgment in the previous revision, and we don't know how to stop changing our minds about p. I take it as a good thing for my claim to be modeling second-guessing (and for my claim to be modeling justified belief) that there is a corresponding potential regress in the re-calibration rule:

$$P_f(q) = P_i[q/(P_i(q) = n \ . \ PR(q/P_i(q) = n) = z)] = z$$

Application of this rule issues in a first-order degree of belief in q. If z did not equal n then this rule applies again to the belief that $P(q) = z$. The fact that z iss my reliability level when I am confident in q to degree n does not imply that z is my reliability level when I am confident to degree z. I will have to go and see what my reliability level is when I am confident in q to degree z. What is to say that this obligation to reconsider does not go on forever, with me never settling on a degree of belief in q? I would not need any new first-order evidence about q in order for there to be this obligation to re-calibrate indefinitely. The process only appeals to the grid of higher-order evidence I (may) have about myself; each step would take a new part of that evidence into account. Re-Cal thus looks like a recommendation for a pathological regress.

Part of the reason we do not consider it healthy to regress is surely a general rule of thumb that says too much thinking is detrimental because it requires resources and postpones decisions. This practical consideration can obviously cut off the indefinite application of Re-Cal just imagined. However, this explanation of stopping would allow that where we stop doubting our judgment is arbitrarily chosen with respect to getting it right about the particular subject matter q. It would be nice if there were an inner logic that yielded a more fine-grained justification for stopping after a particular re-calibration, and that had some relation to coming to a correct judgment of q. After all, we know that there are subject matters where more re-thinking leaves us better off than less would.

There is such a principled way of stopping, which is clear if we understand repeated application of Re-Cal properly. There is obligation and license to re-apply Re-Cal only if second-order evidence is to be taken into account that is distinct from any previous second-order evidence Re-Cal was applied to (concerning q). The Re-Cal rule can be thought of on analogy with Bayesian first-order conditionalization, which is also applied repeatedly as new first-order evidence comes in, and never applied to the same evidence again. We do not consider the fact that the first-order conditionalization rule can be applied again to a distinct set of evidence a regress. We do not worry about the fact that the first-order belief a person has will fluctuate when more evidence is step-wise taken into account in new applications of the rule. The rule is for giving the subject the right degree of belief given a certain set of evidence. Similarly here, it does not make sense to worry about the fact that if I apply Re-Cal to a distinct set of second-order evidence my degree of belief in q might change again. Following Re-Cal, just like following the first-order conditionalization rule, has your first-order degree of belief settle when you have taken all the evidence you have into account, and has it potentially change when new evidence comes to light.[10] Thus, what we have is not a regress but a policy.

---

[10] Of course, it is reassuring (to a Bayesian) that there are theorems showing that if you apply (1st-order) conditionalization to separating evidence to infinity, you will converge to the true probability of q. Re-application of the rule thus would not lead us astray even if we did "regress" by applying it infinitely often. My expectation is that a similar theorem can be proved showing 1st-order conditionalization plus re-calibration would converge in the infinite long run to the truth. This is good, but in turn raises the question what the added value of re-calibration might be.

With this we have the answer about how to choose in a principled way when to stop re-calibrating and go to the next piece of first-order evidence about q: we should stop when our actual higher-order evidence runs out. In pathological judgmental self-doubt, mistakes are made by not recognizing how quickly this happens. The root of this mistake is a sloppy impression that each subsequent stage of doubt asks the same question. Re-calibration is not a matter of asking the same question again and again, but asking at each stage what was the reliability of one's degree of belief in q at the previous stage. The judgment that gives me my degree of belief in q at stage zero is a different process from that combined with a re-calibration to a new degree of belief in q. This means that the content of my judgment of my reliability at coming to my current degree of belief in q is different at each stage of re-calibration, and in turn that the relevant evidence is different.

If we estimate this reliability by track record (which is not necessary but illustrates the point simply) then a different track record is relevant in the second re-calibration than the one used in the first re-calibration. In the case just outlined where I re-calibrated from n to z, a different data set is relevant to judging the quality of my new degree of belief z in q. It is not the whole set of cases in the past where I had initial degree of belief n, and not the set of cases where I had initial degree of belief z, and not the set of cases where I had re-calibrated degree of belief z. The set I need to look at is those cases where I both had initial degree of belief n *and* recalibrated degree of belief z in q. In other words, I have to place my case in the most specific reference class whose properties I have reason to believe are relevant to the outcome and for which I have data. With every new re-calibration that reference class is more specific, which means that the data set available for each further re-calibration will be strictly smaller than that for the previous one because a more specific profile is required for data to qualify as relevant to my case. Even if we had the mental energy for a great deal of repeated higher-order thinking, we would be unlikely to have enough evidence to legitimate it.

Seidenfeld (1985) raises this and other interesting problems for calibration that I discuss in other work, and to all of which I claim there are good replies.

I suspect that in real life that point at which we run out of sufficient evidence to re-calibrate again is reached very early in the process, maybe even typically after the first step, though obviously it will depend on the case. My diagnosis of pathological second-guessing is that by conflating the content of the questions asked at each re-calibration – that is, assuming that the self-doubting question is the same at each stage – one thereby mistakenly thinks one is in a position to make higher-order re-calibration judgments that in fact one does not have the evidence to make. So, philosophy can be beneficial to life in an ironic way: if the second-guesser would think more precisely, and less vaguely, about what question exactly he is asking when he doubts himself, then he would be able to see when it is rational to stop re-thinking.

References

Christensen, David (2007), "Epistemic Self-Respect," *Proceedings of the Aristotelian Society* CVII, Part 3: 319-336.

Dittman, E.T. (2003), "Chronic Second-Guessing Jeopardizes Mental Health," *Monitor on Psychology* (APA) (Volume 34, No. 5 May 2003): 15.

Egan, Andy and Adam Elga (2005): 'I Can't Believe I'm Stupid', *Philosophical Perspectives*, 19(1), 77–93.

Mirels, L., et al. (2002), "Judgmental Self-Doubt: Beliefs about one's Judgmental Prowess," *Personality and Individual Differences* (Vol. 33, No. 5): 741-758.

Seidenfeld, Teddy (1985), "Calibration, Coherence, and Scoring Rules," *Philosophy of Science*, Vol. 52, No. 2, (Jun., 1985), pp. 274-294.

Skyrms, Brian (1980), "Higher-Order Degrees of Belief," in *Prospects for Pragmatism*: *Essays in Memory of F. P. Ramsey*, D.H. Mellor, editor. Cambridge: Cambridge University Press, 109-137.

Appendix

If one still thinks there is an incoherence, it is probably because of the following. If I am certain of the condition of Re-Cal, and also have perfect knowledge of what my beliefs are, the following three things hold:

$P(q) = n$,
$P(P(q) = n) = 1$, and
$P(PR(q/P(q) = n) = z) = 1$

Cal says:
$P[q/(P(q) = n . PR(q/P(q) = n) = z)] = z$

Since the subject's degree of belief that $P(q) = n$ is 1 and her degree of belief that $PR(q/P(q) = n) = z$ is 1, does it not follow that $P[q] = z$? If so, then we have both that $P(q) = n$ (by assumption above), and $P(q) = z$. Contradiction.

It is sufficient for my purposes here that my argument above assume that one lacks perfect confidence about one's beliefs. It is sufficient that $P(P(q) = n)$ does not equal 1 but $1 - \varepsilon$, in which case

$P(P(q) = n . PR(q/P(q) = n) = z) = 1 - \varepsilon$, so

$P(PR(q) = z) = 1 - \varepsilon$

Let $PR(q) = z$ be represented by "B." Then Cal, with its conditional probability rewritten as a ratio, says:

$P(q . B)/P(B) = z$

$P(B) = 1 - \varepsilon$, so

$P(q . B) = z (1 - \varepsilon)$

$P(q) = n$, we assumed, and if q and B are independent, then $P(q . B) = P(q) P(B) = n (1 - \varepsilon)$, yielding

$n (1 - \varepsilon) = z (1 - \varepsilon)$, contradiction.

However, independence of q and B does not come cheaply; $P(B)$ does not equal 1. B is $PR(q) = z$, so assuming that q and B are independent requires assuming that

$P(q/PR(q) = z) = P(q)$

This is more than a violation of PP (which Cal admittedly allows). Violating PP means $P(q/PR(q) = z)$ does not equal z. This is the much stronger claim that your degree of belief in q swings free of what you think the objective probability is; you behave as if objective probabilities are *never* relevant to your degrees of belief. In the argument of the text I am assuming that one is not such a person. If one were then one would have more problems than fallibility. This is an illustration of how easy it is to be misled by considering questions as if they could only involve extreme

probabilities. Extreme probability assignments automatically guarantee independence, and often this is an inappropriate assumption. (Thanks to Jeffrey Dunn for pressing me on this argument that Cal yields incoherence.)