The Re-Calibrating Bayesian

A person is calibrated if his confidence matches his reliability. He is no more strident in his assertion of p than his abilities in figuring out such things would support; he is no more sheepish than his level of fallibility requires. To be calibrated is conceptually distinct from having assimilated your evidence about q in the appropriate way, even if one could be shown to be sufficient to achieve the other under specified conditions. Calibration requires consonance of your confidence in q with general facts about yourself and your circumstances, especially your cognitive abilities, methods, and performance in given types of circumstance, information that is sometimes available in your track record for making such judgments. In the classic example, a weatherman is well calibrated if it rains on 20% of the set of days on which he has 20% confidence that it will rain. We can get a running estimate of whether he is well calibrated by looking at that set of days in the past on which he has expressed 20% confidence in rain, and seeing whether 20% of those days were rainy.

Calibration is a good thing, but what is a rational person to do if she finds herself uncalibrated? It is natural to think that she should re-calibrate, matching her confidence to her newly discovered reliability, and that is the view I will defend here. Natural as it is, this project requires considerable care because common Bayesian assumptions imply that a person who is uncalibrated is ipso facto not rational, or, in other words, that a person must be calibrated, and behave as if she is calibrated, in order to count as rational. Because of these assumptions, the Bayesian framework of rationality cannot give the uncalibrated person any advice at all. The current project is motivated by the thought that lack of calibration is not a failure of rationality, but rather a failure to comport oneself in line with the empirical facts about one's reliability. The role of rationality constraints in such a situation is to tell us how the subject should revise her confidences on learning these empirical facts. The current project effectively provides a generalization of the Bayesian rationality framework.

Track record is not the only way to learn about our reliability and calibration level. Information is also increasingly available from empirical psychology, which studies presumptively average human beings and defined subclasses thereof. The average human being is well calibrated for some kinds of judgments, and poorly for others. In visual perception, for example, arguably the capability most important for our survival, we are extremely well calibrated. We have reliable mechanisms for discerning whether and to what extent in what circumstances our sense organs work properly and we are highly attuned to the cues indicating these states. For example, one normally does not have confident beliefs about what things may or may not exist in front of one if one's visual field is very blurry or black. In those situations we know better than to be confident in any claim that requires current visual information. Our lack of confidence matches our lack of reliability. Normally, in basic visual perception about gross matters, we do not even have to decide how or whether to get ourselves calibrated, or how confident to be. We are equipped not even to consider believing things we are unreliable about.

Things are different with eyewitness testimony identifying individual people as perpetrators of crimes, even though visual perception is involved in this process. In this, psychologists have discovered, human beings tend to be significantly uncalibrated in the direction of overconfidence. Misled by the intensity and vividness of a crime scene experience, for example, we tend to be more sure of who the murderer was than our faculties and positioning justify. Both witnesses and jurors often assume the opposite, that the emotional intensity of the crime scene makes it much less likely for a person to be wrong – how could one ever forget that face? However, the extreme intensity and stress of a crime scene generally makes people even less reliable than normal at reporting the facts, especially unique identifications of faces.[1] To conclude that human beings are unreliable here would be to underdescribe the situation, though. Performance at face recognition varies a good bit with many variables. For example, police officers are not generally found to be better than average people at face recognition, but they are significantly better in situations that more closely resemble the realistic situations they are trained for and face on a regular basis. Reliability has also been shown to improve with intervention on systemic variables, such as how a police line-up is presented to a witness, and may be susceptible to correction after the fact for variables that the police and judicial systems cannot control.

In principle correction on an eyewitness's confidence could be done by a person who is deciding whether to believe him, but here I will be discussing the kind of revision one can do on one's own confidence, and will reserve the word "re-calibration" for this. Calibration is a state. Re-calibration is a process. Intuitively, re-calibrating oneself is adjusting one's confidence in q on discovering information that says one's reliability on q-like matters makes one's current confidence inappropriate. While taking one's evidence concerning q into account can be seen as aiming to get one's confidence in line with the objective probability of q, re-calibration is, in the first place, an effort to get one's confidence in line with one's own reliability about q. These are two different projects that make use of two different kinds of evidence. For example, on witnessing a murder I might become highly confident of the identity of the criminal on the basis of the visual evidence I have about hair color, physique, and facial distinctions. I might, however, subsequently be led to reduce my confidence on learning about the psychological evidence that suggests confident eyewitness testimony is not reliable. The weatherman above might have been uncalibrated. If so, that means that of those days when he has 20% confidence of rain it rains on some percentage not equal to 20. If he learns that it rained on 80% of the set of

[1] In experimental studies, psychologists often measure confidence and *accuracy* – correctness in a particular judgment -- rather than confidence and *reliability* – a tendency to get a particular kind of question right, but the latter is a generalization about the former, and such accuracy data provides the best information in an experimental context for inferring reliability. Psychologists do think they are measuring general trends in how people with particular traits in particular situations subjected to particular procedures do in getting it right.

previous days on which he had 20% confidence in rain, and he adjusts his 20% confidence about rain today to 80%, then he has re-calibrated.

Calibration is generally regarded as good, but re-calibration is controversial, not only because of a worry that individuals may lack sufficient evidence to do it properly, but also for more foundational reasons. The statistician Dawid (1982) argued that on a Bayesian view of rationality and rational updating, a rational subject would not make use of incoming information pertinent to whether he is calibrated or not, but would be constrained simply to assume that he was. This could be seen as something of a reductio ad absurdum of Bayesianism since calibration is a good thing and it is an empirical fact that a person may be uncalibrated at any given time. Thus, it seems to behoove the rational agent to acknowledge that possibility and use what information he has to correct it. However, Seidenfeld (1985) argued that though it was true that the Bayesian subject had to assume he was calibrated, this was just as it should be, since first-order conditionalization alone – that is, properly assimilating your evidence about the original subject matter – leads to calibration in the infinite long run, and in the short-run re-calibration is distorting. There is no point, and much mischief, in re-calibration.

There is also some empirical reason to be suspicious of re-calibration. Isn't that what people do when they second-guess their own judgments? Often people inclined toward this don't know how to stop. Psychologists find that chronic judgmental self-doubt is correlated with debilitating symptoms, such as mood swings, indecisiveness, procrastination, low self-esteem, and anxiety. One could be forgiven for concluding that these people shouldn't have started down that road of free-wheeling self-doubt in the first place. That is, perhaps one should not consider revising one's confidence when one hasn't been given any new evidence about the primary subject matter.

In this paper I will argue that it is possible and good to be a broadly Bayesian subject and also a re-calibrator. The rule for re-calibration that I will formulate and defend is a generalization of first-order Bayesian constraints, and will explain in what sense we are well-calibrated in vision, why and how the eyewitness I described should re-calibrate, and why the chronic second-guesser is not wrong to be inclined to re-calibrate but is rather making mistakes of execution. There are many other applications for a rule of re-calibration. I have argued elsewhere that any pessimistic induction over the history of science requires an assumption that we are obligated to re-calibrate on learning of reason to think we are less reliable than we thought.[2] (Roush 2009) My rule and its defense here explains how and why this is so, while also showing why no similar obligation to lose confidence follows when the Creationist extracts the admission that our scientific theories *might be wrong*.

In the defense of this, much depends on what it means to be a Bayesian and to be a re-calibrating subject. The minimal Bayesianism that I have in mind is personalist: it uses an

[2] Though I assist the pessimist with this part of his argument, I undermine his argument on other grounds, namely a cross induction on method. (Roush 2009)

interpretation of probability in which a statement of probability is a statement of the degree of belief of a subject in a proposition. Thus,

$P(q) = x$

says that the degree of belief of the given subject in the proposition q is x.[3] On this view, a degree of belief is a disposition, "a basis of action," as Frank Ramsey called it, and the disposition can be revealed by the extent of one's preparedness to act on the truth of the proposition believed, for example in the placing of bets. Using a probabilistic representation is not merely a decision to write the matter down using "P's". In writing down the degrees of belief of a subject with "P's" we affirm that the beliefs of this subject conform to the axioms of probability. To be rational, on this view, is for one's degrees of belief to *be* probabilities, whatever else they might be; all of one's x's for all of the q's in one's language – that is, the degrees of confidence one has in each of the propositions of the language – relate to each other as probabilistic coherence, defined by the axioms, requires them to. For example, not only must the subject not believe q when she believes -q – which means she conforms to the consistency constraints of deductive logic – but also her degree of belief in q must be .45 if her degree of belief in -q is .55.

The requirement of conformity to the axioms is weaker than it is often taken to be, and in a way that is especially relevant here. A locution that has the subject "assigning probabilities" is often used interchangeably with that of the subject "having degrees of belief." However, since in personalist Bayesianism a probability *is* a degree of belief these cannot be equivalent, because for the subject to assign probabilities would then be for him to act directly upon his beliefs to determine them. This wouldn't be possible since belief is not voluntary, but it is also, of course, not what is meant by "assigning probabilities," where the picture is that the subject chooses a number *indicating* how likely he thinks an event is. This reporting or designation of one's degree of belief may of course occur, and even be helpful, but it cannot be what a probability *is* in the personalist interpretation; a subject need not do a mental act of choosing, thinking about, reporting, or even understanding the concept of, a probability in order to *have* a degree of belief. Second-order probabilities also do not require any potential awareness on the part of the subject of what his opinions about his first-order degrees of belief are, for the second-order beliefs are also beliefs. A subject's having a degree of belief corresponds to his having a disposition to act, for example to bet, and his having a degree of belief about his degree of belief corresponds to his having a disposition to act, for example to bet, on what his degree of belief is one level down, but

[3] This kind of subjective Bayesianism thus does not fall prey to the familiar objections that ordinary people don't assign probabilities, and that probability can't be a model for understanding scientific inference since there were many rational scientists before the concept of probability was even invented. Such objections are not to the point, since presumably people do have degrees of confidence. Those need not be exact either, in order for the Bayesian model to be a good idealization and to yield illuminating qualitative and ordinal relationships.

neither requires reporting or awareness, even potentially, of either belief.[4] In an experiment we could ask him how he would bet on what his bet would be on q, without any reference to his beliefs. This elicitation does not even require knowing that in betting a particular way one is revealing one's degree of belief. To be probabilistically coherent a subject's beliefs must be related in certain ways, but he can be immune to Dutch booking without awareness that he is, and without any deliberate self-guidance to this end.

These distinctions are important here since sloppiness about the difference between beliefs and beliefs about one's beliefs, the relation of belief to probability, and the role or lack of role for awareness and acts of assignment, can lead to false conclusions and obscure possibilities. For example, the weatherman both simply *has* degrees of confidence in rain and, in considering whether he is calibrated, would typically consciously consider properties of his beliefs. If he appears uncalibrated, he might come to a different degree of belief about rain today in light of this information. He probably would also report probabilities, translating his confidences into statements of objective or subjective probability, or vice versa. This involves degrees of belief and reports of probabilities explicitly. By contrast, what would happen were your visual field to become entirely black is that you would cease to have confidence in claims about objects of visual perception that required ongoing evidence, and you wouldn't have had to think at all or be able to report anything to yourself or others in order to achieve that. Both are clearly re-calibrations of first-order degrees of belief on the basis of information about the subject's own reliability, but one case involves awareness and reports of probabilities and the other does not involve even potential awareness. If one thought that having beliefs and second-order beliefs – beliefs about one's beliefs – required potential awareness of or acts upon one's beliefs, one would have a hard time making out what the similarity here is. Carefully abiding by a personalist Bayesian view will allow us to see that what is essential to re-calibration, just like what is essential to coherence, is how various degrees of belief should respond to changes in other degrees of belief, not what mechanisms or acts enable a particular subject to achieve those responsiveness relations. With respect to the end of re-calibration the means are a contingent matter.

The minimal Bayesian requirement is also stronger than it may seem. We believe lots of things, and who among us is consistent, as probability requires? We do not have degrees of belief for every proposition of our language. We should not be perfectly confident about every logical truth and falsehood as to which is which, since for some of them the jury is still out among the most sophisticated set theorists, yet a typical probabilistic representation requires these things. I will ignore these mismatches here, partly because they all are consequences of the fact that Bayesianism is an idealization. Idealizations play a fruitful role in inquiry, leading to

[4] Awareness and knowledge are not equated here. The current point is that one may have degrees of belief without awareness of them, but some, including this author, think one can have knowledge of p without fulfilling any awareness requirement. It requires a distinct, further argument, given below, that one may be a rational agent yet not have knowledge of what one's beliefs are in virtue of those beliefs about one's beliefs failing to be true.

explanations, discoveries, and questions. One thing that their fruitful use teaches us is that there may never be a day when we have one model that captures just every aspect of a topic. To cry about this would be kind of like crying about life in general though, briefly satisfying but generally unproductive.

It is good for an idealization to be simple; that makes it tractable when the subject matter construed more realistically is not. Also for reasons of tractability, the fruitful way to improve an idealization is one step at a time, which is what I set out to do in this paper. It just happens that the areas listed above where Bayesianism does not model our epistemic lives very realistically are not the idealized aspects I have chosen to discuss here. Since relaxing idealizations makes a representation more realistic, it tends also to make it more complicated. However, I still aim to achieve the greatest increase in explanatory power with the least added complication.

Personal re-calibration and second-order beliefs

Defending re-calibration requires a precise representation of what it is. Many authors discuss calibration using first-order probabilities, that is, degrees of belief about matters that do not involve degrees of belief. This is sensible for describing the calibration state of another subject, but the first thing I will argue is that if we use probability at all to model personal re-calibration, then the use of second-order probabilities – probabilities of probabilities – is not only useful but required. This is because re-calibration involves revising degrees of belief on the basis of degrees of belief about properties of degrees of belief, and degrees of belief are probabilities. To ignore this structure would result in a misleading underdescription.

It is not uncommon to hear the protest that second-order probabilities are too complicated to fathom. However, some epistemologists are quite comfortable talking about second-order beliefs, and appealing to intuitions about them, while using probability to model first-order beliefs. Intuition is also used to decide how a given second-order belief should affect the first-order probabilities. Because the relation between first- and second-order probabilities involves delicate issues and requires choice of a rule of relation between the levels, using intuitions in individual cases amounts to helping oneself to a powerful free parameter. Beliefs about beliefs *are* probabilities for anyone using probability to model belief. Thus, one's options in this area are 1) not to speak of second-order beliefs at all, 2) not to use probability to model either second-order or first-order beliefs or 3) to use probability at both orders.

An example will illustrate the fact that second-order structure has an ineliminable role in re-calibration. Imagine two fields of vision, one filled with a leafy, jungly scene and lacking indicators of tigers, the other entirely black, and thus also lacking indicators of tigers. The subject possessing the first field has more information than does the second subject concerning whether a tiger is present. Arguably, the first subject also has a different level of justified belief that there is no tiger; he should be relatively confident that there isn't one, while the second subject should not. Yet the evidence their visual fields have concerning *tigers* is the same. Neither of them has indicators of tigers; neither has percepts of stealthily moving orange and white stripes, for example. Do they differ in their evidence about absence of tigers? Depending on how we like to use the words, we might say that neither has indicators of an absence of tiger within the visual field or we might say that both have indicators of absence in all those pixels that do not exhibit the characteristic orange and white stripes. Either way, the information within the visual field that concerns *tigers* does not break the symmetry of the information available to these two subjects. To explain their very different epistemological situations we have to consider their evidence about their evidence. The black visual field is an indicator, to a normal subject, of the fact *that he has no visual evidence* of whether there is a tiger or not, that a belief of no tiger that was formed on the basis of beliefs about those pixels would not be trustworthy. His appreciation that the total blackness of the field is an indicator of his unreliability is a second-order fact, a belief about his visual-field beliefs or evidence. The tiger case illustrates another thing about evidence useful for re-calibration: it need not take the form of a track record. It is possible to possess a faculty that gives us concurrent and generally true feedback on itself, and it appears that we have just such a thing in vision.

In re-calibrating a confidence about q, the information we use is not about q per se but about reliability, which necessarily brings in beliefs about beliefs. What we have just seen is not only that there must be beliefs about beliefs in any model of re-calibration, but also that they must be beliefs about the subject's *own* beliefs. In re-calibrating we are not per se concerned about the reliability of other subjects but of ourselves. The reliability of others may be relevant to mine insofar as I am depending on them for forming my confidence about q, but then they are part of my mechanism for forming belief and their contribution is, or should be, taken into account when I evaluate my reliability. To re-calibrate, the beliefs I need to have beliefs about are my own.

To develop a language for re-calibration, we begin by describing a subject's belief about q using first-order probability. Thus, I write about subject S:

$$P_S(q) = x$$

which means that S has x degree of confidence in q. I can describe S's reliability as an objective probability (of whatever sort one likes) using a probability function I will call "*PR*." Thus, S is reliable to degree y when believing q to degree x iff:

$$PR(q/P_S(q) = x) = y$$

which says that the objective probability of q given that the subject S has degree of belief x in q is y. *PR*, though a probability function, is a different function from P_S, and is not interpreted as degree of belief. Thus, I am not yet representing second-order degrees of belief. *PR* may be chance, frequency, propensity, or whatever objective interpretation one prefers. The account of recalibration does not depend on this. Typical calibration curves reported in empirical psychology justify the specificity of reliability level to the degree of belief one has in q. In many domains we have different levels of miscalibration at different levels of confidence, often being overconfident when confident and underconfident when lacking confidence.

I can also describe the state of S's being calibrated in one's degree of belief x:

$x = PR(q/P_S(q) = x)$

which says that when S is confident to degree x about q, her degree of belief in q is x, i.e., matches her reliability about q at that confidence. One could represent a subject as being calibrated for q full stop by adding a universal generalization over x:

$(\forall x)\ x = PR(q/P_S(q) = x)$

S's beliefs about q, and their reliability properties, can be faithfully described by us without any nesting of a subjective probability function within a subjective probability function. First-order probability is sufficient for discussing the calibration state of a person who is not oneself.

We can describe a situation where someone else has beliefs about S's beliefs, by nesting the foregoing statements in a subjective probability function different from S's, the function that represents the degrees of belief of T:

$P_T(P_S(q) = x) = z$

This says that subject T believes to degree z that S believes q to degree x. Similarly, we can describe T's belief about S's reliability:

$P_T(PR(q/P_S(q) = x) = y) = z'$

which says that T believes to degree z' that the objective probability of q when S believes it to degree x is y. Intuitively, if Tonya believes to degree .95 that the objective probability of q when Sam believes it to degree .9 is .5, this means that when Sam tells Tonya confidently that q, she behaves as if he has not given her any information whether q. This would provide a gross model of the response of a juror to an eyewitness whom she regards as having no credibility at all. In that situation we are imagining that Tonya has a very precise view that the witness's lack of calibration on q is in the direction of overconfidence, and to a degree that makes his beliefs exactly useless. We describe the situation where T is highly confident only of the weaker claim simply that S is not calibrated by writing:

$P_T(P_S(q) \neq PR(q/P_S(q) = x)) = .99$

This says that T is highly confident that S's confidence about q (his degree of belief in q) does not match S's reliability about q (the objective probability of q when he has that confidence).

We have represented one person's beliefs about another person's beliefs using two subjective probability functions, one for each person, and nesting them. To represent a person's beliefs about her own beliefs I will use a single function nested on itself. The expressions we have already used, such as:

$P_T(P_S(q) = x) = z$

present complications, giving us the need for work like Gaifman's (1980) theory of higher-order probability to make sense of them. However, there are even more challenges posed by the special case where we let P_T and P_S be the same function. I think that these added challenges should be expected in modeling our phenomenon, given that we are dealing with the beliefs of one person and, intuitively, judgmental self-doubt looks close to inconsistency. A person worried about her beliefs is definitely in conflict with, and in that way interacting with, herself. Nevertheless, she is also still one person, not two. If we represent a belief about one's own belief by nesting a single probability function around itself, then the nesting that allows two different orders is how the subject's inner conflict can be displayed, and the use of a single probability function will be part of how the unity of the subject is retained. I will defend the coherence of this picture in what follows.

We will represent these things as modifications of the previous equations:

$P_S(P_S(q) = x) = z$

This says that subject S believes to degree z that she believes q to degree x. Similarly, we can describe S's belief about her reliability:

$P_S(PR(q/P_S(q) = x) = y) = z'$

This says that S believes to degree z' that the objective probability of q when she believes it to degree x is y. We describe a situation where S is highly confident that she is not calibrated, assuming her degree of belief in q is x, by writing:

$P_S(P_S(q) \neq PR(q/P_S(q) = x)) = .95$

S believes to degree .95 that her degree of belief in q is not equal to the objective probability of q when she believes it to degree x.

Notice that these equations are exactly the same as the previous ones concerning T's beliefs about S, only with "P_S" substituted for "P_T". We are representing a subject as taking with respect to herself a point of view that is as external as, and the same as, any other person would be forced to take when provided the same information about what S's belief is and about S's reliability. Yet because it is the same function providing this view as provides S's first-order beliefs, this external view of herself is also as much her own view as her belief that the sun will rise tomorrow is. Although the subject has inner conflict she remains one subject; she is simply a person with an epistemic worry, seeking self-improvement.

Use of a single function imposes a certain unity, but it is not the only thing needed to hold a probabilistic subject together. She must maintain coherence, of course, and as we should expect intuitively when modeling a subject who is doubting her own judgment it will be challenging to understand how this is possible. Moreover, the minimalist Bayesianism described above does not dictate the relation between the two orders; it is easy to see syntactically that the axioms give constraints only within an order, not between orders. Thus, any bridge principle that may be adopted between these two orders constitutes an independent axiom.

In the Bayesian literature so far, the issues about how the two orders of probability should relate that are relevant to rational self-doubt have been concealed from attention by idealizing assumptions, as I will explain. Motivated by the empirical observation that it can be rational to doubt oneself and to revise one's original belief on that basis, I will generalize away from those assumptions. The biggest challenge will be to explain how the self-doubting subject who can be represented in this generalized framework could be probabilistically coherent. But there is an intuitive question corresponding to this as well: Is it possible to cope with an incident of self-doubt without either becoming just a heap of parts, or exiting the state by instinctive fiat? Can we learn in an orderly fashion from the things that prompt self-doubt?

Second-Order Probabilities

Even many of the greatest defenders of probabilistic rationality constraints have had some resistance to second-order probabilities, regarding them as suspicious when not trivial. (de Finetti, Savage, Good, Jaynes, Levy, Seidenfeld) I have just argued that they are necessary for a proper analysis of self-doubt, but it remains to show that they are possible, that is, coherent, especially in the extreme self-referential form I am advocating.

Classic objections, which are still often heard, were elegantly addressed by Skyrms (1980). For example, one might think that second-order probabilities are well-defined, but useless because trivial. They will all be zeros and ones, appropriately distributed, because the rational subject should be certain of what her beliefs are and are not, and should be right about them. Such extreme probability values can neither change nor effect a change in any other proposition's probability. They are thereby trivial because inert. One might think these probability values should be zeros and ones because of a picture in which introspection of one's mental states is special and infallible.. However, even among those who think introspection is distinctive and of crucial importance in epistemology this infallibility assumption has long been discredited. In contrast to the infallibility assumption, one might take a dim view of our introspective capacities but nevertheless fall into a similar trap, thinking that even first-order degrees of belief don't exist *because* we can't introspect them. Introspective access to what our beliefs are, or indeed any kind of infallible knowledge, is not a precondition of their existence on Ramsey's view of beliefs as dispositions to act.

Others have presented a conundrum for the betting method of determining someone's degrees of belief: if we had a subject bet on what her degrees of belief are, then she would have an incentive to bet misleadingly at the first-order to protect those initial bets. However, not only should we have more confidence than that in experimenters' ingenuity,[5] but also, a belief, a disposition to act, is not just the same thing as its method of verification. We could think of the introspective access and verificationist objections as manifestations of right-wing and left-wing positivism, respectively. (Skyrms 1980)

An advocate of the idea that second-order probabilities should be zero and one still has a plausible reply, it seems to me. Beliefs are dispositions to act, and we know very well that we are not always perfectly acquainted with those. We sometimes become acquainted only when we witness ourselves acting, he admits. But the probabilistic conception of rationality is a normative one, and we do not have to suppose that as a matter of fact we are infallible about these things in order to assume that we would be ideally rational if we were. However, while it is true that it is thus logically consistent to require infallibility about our beliefs while admitting we do not have that, this picture is inconsonant with a probabilistic idea of rationality in another way. Bayesian rationality puts constraints on the relations of one's substantive beliefs to one another, but does not take one to be obliged to have accurate degrees of belief about empirical, or more broadly substantive, matters. Someone who has false beliefs about the laws of nature, who stole the cookie, or the population of his county, is mistaken, but not thereby irrational, on this kind of view. There are more substantive conceptions of rationality, but indeed those who favor them lament the fact that Bayesianism puts no constraints on the subject's priors. It cannot be denied that whether I have a certain degree of belief in q or not is an empirical matter, and it would be exceptional for Bayesian rationality to require perfect knowledge of such a thing.

There is a possible reply to this line too, I think, which is that the stipulation and special treatment is necessary because the self must be seen as having a special relation to its own beliefs in order to *be* a self. It would otherwise, in one way or another, be a heap, disunified, dysfunctional, incoherent. This is an objection that will require much of the remainder of this paper to address fully. It does appear, as I have said, that having a relation to oneself that others do not bear to one is essential to being a self. However, the bridge principles that I will defend below give one a relation to oneself that others do not have, without requiring perfect, or even good, knowledge in order to achieve this. We will see that it is neither self-knowledge nor self-respect, but rather the disposition to do the right thing in response to one's imperfections that insures the epistemic unity of the self.

[5] The relative sizes of the bets can be adjusted to minimize this distortion.

It is relatively easy to fall into equivocations that lead to the impression that second-order probabilities involve contradictions. David Miller (1966) presented an apparent paradox that involved a conflation of de re and de dicto readings of probabilities:

1. P(-q) = P(q/P(q) = P(-q))

2. P(q/P(q) = P(-q)) = .5

Therefore, P(-q) = .5

q was arbitrary, so since it can't be that every proposition has 50% probability we have an inconsistency and probabilistic incoherence. The problem, as Skyrms (1980) pointed out, is two possible readings of "P(q) = P(-q)." The assumption that sneaks in the particular designation .5 as the probability of the arbitrary –q is a de re reading of the embedded "P(-q)" and a de dicto reading of the "P(q)" in the first premise, whereas in the second premise "P(q) = P(-q)" is read de dicto. If P(q) = P(-q) then P(-q) does equal .5 But if P(q) as a matter of fact is .75, then P(q) ≠ P(-q). The probability of P(q) = P(-q) need not be zero for this to be so, so the first premise could be defined and false. It would be false for every value of P(-q) except .5. That is, Miller's argument would be unsound except in those cases where the conclusion advertised was true.

Skyrms thereby defended the legitimacy and coherence of a useful bridge principle between first and second order probabilities:

$P_2[q/P_1(q) = a] = a$

He called it "Miller's Principle" in honor of Miller's contribution to its discovery. Assuming that both P_1 and P_2 belong to a single subject, the principle says that (if he is rational) his degree of belief in q given that his degree of belief in q is a will be a. There is nothing incoherent here because the "a" rigidly designates a particular number. He pointed out that we can generalize to make "a" a variable, say "x," as long as we do so uniformly. This is a principle that I will generalize in order to model rational self-doubt.

Skyrms uses two different probability functions for the first and second order. I will use the same probability function for both. Miller used one function in his argument for a paradox, but a problem elsewhere in the argument was sufficient to avoid that paradox, so we do not yet have a problem for my view. Nevertheless Skyrms chose a typed theory – with a different function at each order – to avoid a different problem that I must now address. Consider the collection of propositions, and imagine its power set, the set of all of its subsets. As is clear intuitively, we should not expect to be able to map the power set of a set into itself. If a set has two members, 1 and 2, the power set has three, {1}, {2}, and {1, 2}. This generalizes; the power set is always strictly bigger than the set it is power set of. However, if we allowed a probability function to apply to its own probability statements as propositions, we could produce the impossible mapping from the power set into the necessarily smaller set itself. Let S_1, S_2, S_3, …

be the subsets of the set of propositions. For each one, we can construct a proposition about it. E.g., George believes p if and only if p is a member of S_1. Since all of the subsets are distinct from each other, each of these propositions is distinct. "George believes p" is itself a statement of probability, and George's probability function, which applies to itself, has values for these propositions too because George has beliefs about them. Thus, we have a mapping from the power set of the set of propositions into the set of propositions. Contradiction.

Never allowing a probability function to apply to its own statements – hence using a new function at each new order – prevents this problem analogously to the way that forbidding any statement that a set is or is not a member of itself avoids the Russell Paradox. Using P_1, …, P_2, …, P_3, … assures that there is no one mapping that gives a value for all of the propositions about subsets that could be defined; those defined at one level can only be represented as believed by using a new probability function. Typing the theory will thus rescue it from incoherence. However, we know that the set theoretic paradoxes have more than one possible way out. For example, we can run the foregoing argument as a modus tollens on the assumption that the collection of propositions is a set. We can, and I will, regard it as a proper class, that is, a class that is not a set. Since it is not a set this power set argument does not apply. There should be no objection to this from the intuitive side either. It is precisely because the statement about George's beliefs looks as good as any other as a proposition that we were able to generate the paradox. A collection of things that has such a generative capacity is not a set, we could say, because it does not behave as a set should. It is fortunate for me, from the technical side, that
Herman Rubin (1969)
has shown how to modify the Kolmogorov probability axioms so that probability functions can take proper classes as their domains, through an axiomatization that appears to be superior on other grounds as well.[6] Thus the class of propositions being a proper class need not bring contradictions when we apply probability functions on it.

There are further coherence challenges for the framework I am developing here, as we will see below, but those are specific to the framework rather than objections to second-order probabilities in general.

Miller's Principle and Epistemic Self-Respect

There is more than one variation on and interpretation of Miller's Principle. One descendant is Bas van Fraassen's Reflection Principle which says, roughly, that you should believe what you think your future or present self believes. In Gaifman's (1980) interpretation of his synchronic

[6] A new approach to the foundations of probability. 1969 *Foundations of Mathematics (Symposium Commemorating Kurt Gödel, Columbus, Ohio, 1966)* pp. 46--50 *Springer, New York*

version of Miller's principle, the first-order probability function represents the state of maximum possible knowledge, which may or may not be perfect knowledge, while the function nested around that one is the probability function for my subjective degrees of belief. In this case the principle says that I should believe to the degree that I think the maximal knower would believe. While important, these formulations would not be apt for my question here. Skyrms's formulation

$P_2[q/P_1(q) = x] = x$ MP

involving two subjective probability functions belonging to the same subject, and providing for the possibility of a learning principle, is the appropriate starting point. The domain of the function P_1 is first-order propositions, and the domain of P_2 is first-order propositions plus propositions about P_1's values for first-order propositions. MP requires that the values the function P_2 has for first-order propositions be the same as those it regards P_1 as having.[7] This naturally led to a special case of second-order conditionalization that was equivalent to first-order Jeffrey conditioning. (Jeffrey XXXX) Second-order conditionalization is well-defined, and makes possible a defense of Jeffrey learning via an argument about conditional bets.

MP could be seen as a requirement for deference of the second-order self to what she believes the first-order self's degrees of belief are or have come to be. The second-order self's degree of belief in q given that the first-order self believes it to degree x should be x. If we adopt this principle it is clear that judgmental self-doubt will not be counted rational. Nothing the second-order self might learn could be taken to give reason to think something different than the outcome of the first-order conditionalization that she learns about. Moreover, even if the second-order self had grounds for disapproving of the first-order self's opinions, she would have no way to teach or enforce it on the first-order self, since the first-order self does not in turn condition on the opinions of the second-order self.

Though Miller's Principle is true and illuminating for a wide range of cases, Skyrms (1980, 125) himself pointed out that it would not hold in cases where one believed that the way one came to one's degrees of belief about first-order matters was biased. It would not be rational for the subject to allow the verdict of a process she believed to be so tainted to stand, as MP requires. This is exactly the kind of case discoveries of evidence for our unreliability or lack of calibration require us to think about. MP must be relaxed to account for these cases in a sensible way. The probabilities P_2 must "compensate" for the projected bias in P_1, as Skyrms put it.

[7] Miller's Principle appears to be incompatible with P_2 having inaccurate beliefs about P_1's beliefs. This rule only explicitly requires a relation between P_2's degree of belief in q and the degree of belief in q that P_2 thinks P_1 has. However, if P_2 does not have perfect knowledge of P_1's beliefs, then P_2 and P_1 could have different values for q. Which one of these functions would answer when we asked the subject to bet on q? It appears the only way to avoid this indeterminacy is to require P_2 to be perfectly accurate about P_1's degrees of belief. Thus, Skyrms's MP appears to undermine the fallibility he wanted for 2nd-order probability, a fallibility that had also secured the non-triviality of higher-order beliefs. This is another reason to generalize MP.

If we intend to formulate a principle of conditionalization for how one should update on discovering both that one has a certain degree of belief and that one suspects serious bias in it, then it turns out that separate functions for the two orders will not do the job. If the subject believes q to some degree and also has good evidence that she is overconfident, the rule for properly handling the situation should ultimately lead to her revising her degree of belief in q. If we modify MP in the natural way for this case so as to indicate that P_2 doesn't approve of P_1's belief, we would replace the second "x" with "x – y," $y \neq 0$:

$P_2[q/P_1(q) = x . r] = x - y$ MP'

where "r" refers to whatever evidence P_2 has for this disapproval, and "y" is a discount applied to the subject's confidence in q. Though this formula would underwrite a conditionalization in which P_2 comes to have value x - y for q on learning (r and that) P_1's value is x, it would not require any corresponding change in the value P_1 assigns to q. P_2 could disapprove of P_1's values, but it is not by this formula licensed to intervene on them. If the functions P_1 and P_2 thus have different values for q, there is also a problem of indeterminacy; when we ask the subject represented by these two functions how much she is willing to bet on q, which of these functions should we expect to get the answer from?

In this way we can see that this first revision of MP does not allow self-doubt to be rational. Much less does it allow us rationally to resolve such a situation. To allow the second-order self efficacy in revising the first-order self's beliefs, we must represent both selves by the same function. The difference between the two orders is thereby represented not by the existence of distinct functions, but by the fact that "P" may be alone or may be applied to itself. A rough version of such a formula would be:

$P(q/P(q) = x . r) = x - y$

Your degree of belief in q given that you believe q to degree x and that you also have evidence r that you are overconfident by an amount y, should be x - y. As a diachronic principle,

$P_f(q) = P_i(q/P_i(q) = x . r) = x - y$

this would say that when one finds that one believes q to degree x and one also has evidence r saying one is overconfident by an amount y, then one should come to have degree of belief x – y in q. The self-monitoring the subject represented by P is doing has the chance to lead to self-correction, unlike in the case where we used two different functions, because in using only one function we insure that the function being monitored is the same as the function that is taking a different value in response to that monitoring.

We have done two revisions of Miller's Principle here, one in which we restricted attention to the special case where the subject has only one probability function, that is, where P_1 and P_2 are the same, and the other in the direction of generalizing to the case where the "monitoring" subject need not approve of the deliverance of a first-order conditionalization. However, there are still cases in line with MP in the second respect, where the monitor has no right to disagree, namely, those cases in which she has no reason at all to think her verdict on q was flawed. As we should expect, this case is present as a special case in the new formulation that uses just one probability function. We say:

P(q/P(q) = x) = x *provided (the condition does not have probability 0 and) there are no statements of probability, in the condition or the background, for which* P *has values and that when conjoined with* "P(q) = x" *are probabilistically relevant to* q. (RSR)

Your degree of belief in q given that your degree of belief in q is x – and (roughly) nothing else relevant – should be x. In other work I have called this principle "Restricted Self-Respect." In its diachronic form one could think of it as saying that the mere discovery *that* you have a degree of belief does not provide a reason to change it. A self-doubt that violated this principle would not be defensible.

Intuitively, rational self-doubt needs a reason, and the reason does not come in the form of new first-order evidence, evidence about q; first-order conditionalization will anyway tell the subject what to do with that. If I am confident that the murderer I saw is guy number 2 in the sequence of pictures but then I suddenly remember that the murderer had an earring, there need be no *self*-doubt. If I am a responsible subject the new earring-memory will make me reconsider whether guy number 2 is the murderer, but this should take care of it, unless the tardiness of the recollection produces general self-doubt about my beliefs that come from memory. The kind of evidence that leads us rationally to doubt our own judgment takes a different form from first-order evidence. It refers to our own beliefs and makes probabilistic claims about them, for example: confident eyewitnesses tend to be overconfident, and I am a confident eyewitness. When we take such evidence to heart and doubt our judgment we are not violating the restricted principle of self-respect (RSR) – which says not to doubt yourself without a reason – but only a much stronger principle:

P(q/P(q) = x . r) = x (*provided the condition is defined) and regardless of any other statements of probability for which P has values, whether in r or in the background.* (USR)

This principle, which I will call "Unrestricted Self-Respect," says that your degree of belief in q given that you believe it to degree x should be x *regardless of what else you believe*. Regardless of whether an expert tells you you are not fit to judge, regardless of whether you know that you are on hallucinatory drugs, etc.

Intuition says that unconditional respect for one's own opinions is not sensible, but there is a stronger argument than intuition for rejecting this principle. We can see what is at stake here probabilistically by representing the kind of second-order evidence in question, which I have so far labeled "r," explicitly. r is that statement of the subject's reliability discussed above, and representing that explicitly will both yield the new rule and justify rejection of USR.[8] The claim that a subject has reliability level z when believing q to degree n is written:

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

which is read "The objective probability of q given that the subject believes q to degree n is z." The kinds of discoveries psychologists have made about the unreliability of eyewitnesses would usually have implications that take just this form for the individual whose function is P.[9] If the subject does have degree of belief n, and n does not equal z, then she is uncalibrated. If you are this subject, then your acknowledgement of your belief and of the psychologists' findings leads you to the following conjunction of beliefs:

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

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. If n does not equal z, and the concept of calibration is in your vocabulary, then you should believe that you are not calibrated. The question what this 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:

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

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 seems plain wrong intuitively, and a version of the familiar Principal Principle will explain why. Notice that with a natural assumption,[10] the conjuncts of the condition in !:

$\mathbf{P(q) = n} \,.\, PR(q/P(q) = n) = z$

together imply an objective probability for q:

$PR(q) = z$

[8] I am using USR in this argument for ease of presentation. The argument can be adapted to justify rejection of Christensen's SR (Christensen 2007) because "$PR(q/P(q) = n) = z$" can be represented as in the condition or in the background with probability 1 indifferently.

[9] The psychologists' results are about human beings in general, the average human being. It is possible that a given subject has further evidence showing that she is not average in some way that makes a difference to this reliability issue. However, without such further evidence the narrowest reference class she can put herself in is "human being," and she must assume she has the properties that class is known to have.

[10] $P(PR(P(q) = x) = 1/P(q) = x) = 1$, 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.

so ! should have the same value as the expression

$P(q/\mathit{PR}(q) = z)$

The Principal Principle says[11]:

$$P(q/\mathit{PR}(q) = z \,.\, r) = z\,, \qquad \text{(PP)}$$

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:

$P(q/\mathit{PR}(q) = z) = z$

implying that

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

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

To decide what is rational when n does not equal z, we are forced to choose between the Principal Principle and Unrestricted Self-Respect. PP is less fishy, and it also explains our intuitions about taking information about one's reliability into account, whereas USR conflicts with them. PP is false with inadmissible r, but we have not been appealing to anything intuitively inadmissible. Thus, I advocate rejecting Unrestricted SR, while maintaining PP and 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 conditions (on the natural assumption in fn.10 and assuming that one fails to have either perfect confidence or perfect accuracy about one's degree of belief in q or perfect confidence about one's reliability, a matter I will discuss presently):

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

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

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

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. We can see what the end state of that updating looks like by noticing that the conjunction in the condition:

$P_i(q) = n \,.\, \mathit{PR}(q/P_i(q) = n) = z$

implies

[11] This is more general than the usual Principal Principle in virtue of its taking any kind of objective probability, rather than only using chance.

$PR(q) = z$

Thus, on applying Re-Cal you have:

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

Or

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

That is, you are now back in line with the Principal Principle. Notice that this is the state I labeled above as *calibration*. The forced choice we had here between USR and the Principal Principle and the fact that choosing the latter gave us Re-Cal, appears to imply that rejecting Cal and Re-Cal requires rejecting the Principal Principle as a synchronic and short-run diachronic constraint on the rational subject.

The relation that Cal and Re-Cal have to the Principal Principle also makes the issue of whether the former are coherent even more acute. Since given some reasonable assumptions PP implies Cal and Re-Cal, if they are incoherent then PP is too. If PP *is* a coherent constraint then any incoherence there may be in Re-Cal or Cal comes from the reasonable assumptions made above. One of these was a denial of perfect self-knowledge, and this turns out to be an important issue for the coherence of Cal and Re-Cal.

Self-Knowledge and Coherence

The key to the coherence of Cal (and Re-Cal) is located in the assumptions implicit in my derivation of a forced choice between Unrestricted Self-Respect and the Principal Principle. That choice depended on how we answered the question:

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

which in turn depends on this question having an answer at all; that in turn depends on the condition being coherent. Intuitively the question makes sense – what should my degree of confidence be in q if I believe my confidence is n, but I also believe that the objective probability when it is n is z? However, for the question of the formula to be defined appears to require that being certain of its condition is a probabilistically coherent state to be in.[12] Andy Egan and Adam Elga have argued that one cannot probabilistically coherently maintain high confidence in q and also believe that one is unreliable about q. This would seem to correspond to having high n and low z in our expression, and the condition in our formula allows for this. However, I have argued elsewhere (2009) that their analysis does not settle the issue of whether one can coherently maintain confidence in q while also attributing to oneself low reliability and having decent self-knowledge, because they do not express the questions explicitly using second-order

[12] If one only ever did Jeffrey conditionalization then the condition would not need to be coherent when assigned degree of belief 1 since one would never need to have degree of belief 1 in the condition.

probabilities. Probabilistic coherence alone only puts constraints within an order, not between orders, e.g. first- and second-order. To settle between-order rationality questions one needs explicitly to consider bridge principles that are independent of the probability axioms.

The condition in question in my analysis is a potential source of incoherence via the following argument:

If I am certain of the condition of Cal and also have perfect knowledge of what my beliefs are, the following three things hold:

1. $P(q) = n$,
2. $P(P(q) = n) = 1$, and
3. $P(\mathit{PR}(q/P(q) = n) = z) = 1$

2. and 3. together express my certainty in the condition of Cal, but 2. also expresses one part of the perfect self-knowledge we are concerned with. 2. expresses my perfect confidence that my degree of belief is n. 1. and 2. together express the other part of that perfect knowledge, namely my perfect accuracy: my degree of belief is exactly what I am perfectly sure it is.

Cal says:

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

Since the subject's degree of belief that $P(q) = n$ is 1 and her degree of belief that $\mathit{PR}(q/P(q) = n) = z$ is 1, it follows that $P[q] = z$. If so, then we have both that $P(q) = n$ (by assumption above), and $P(q) = z$. Contradiction. This shows that I cannot both coherently abide by Cal and Re-Cal and have perfect self-knowledge of my beliefs and perfect confidence about my reliability.

If we deny perfect self-knowledge, by assuming non-extreme degrees of belief that $P(q) = n$ or that $\mathit{PR}(q/P(q) = n) = z$, or by assuming that $P(q)$ is not quite n, then this argument does not go through. Since one may make these confidences just as high as one likes as long as they are not 1 and still avoid the incoherence, I conclude that the impression of incoherence is an artifact of the common and hazardous assumption of extreme probabilities.[13]

[13] Failure of the extreme property of Confidence does not prevent one having high confidence about one's degree of belief. It may equal $1 – \varepsilon$. In this case:

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

(assuming for illustration that one is perfectly confident in the second). Let $P(q) = n \,.\, \mathit{PR}(q/P(q) = n) = z$ 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. If we further assume that q and B are independent, then

One might have the idea, suggested above in the luminaries' resistance to second-order probabilities, that the subject would fall apart if she didn't have this self-knowledge, but her not having this perfect knowledge blocks an argument that says Cal and Re-Cal are incoherent, and Cal and Re-Cal are precisely the bridge principles that can hold the rational subject together when she lacks perfect knowledge of herself.

(Thanks to Jeffrey Dunn for pressing me on this argument that Cal yields incoherence.)

The crucial step in the incoherence argument against Cal was securing independence between *PR*(q) = z and q. The perfect self-knowledge assumption that led us to this independence is unrealistic, and the extreme probabilities that we expressed that assumption in imported artifacts. However, there is also a more general argument against the independence that is needed for the argument: Assuming that q and *PR*(q) = z 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; since z and q were arbitrary, this says you behave as if the objective probabilities are *never* relevant to your degrees of belief. In the framework developed here, I am assuming that one is not such a person. If one were, then one would have more problems than fallibility.

There remains an implication that may seem strange: a subject who *does* have perfect knowledge of her beliefs and perfect confidence about what her reliability is must not re-calibrate via Re-Cal on pain of incoherence. This gives us two options, one of which is to make an exception to Re-Cal, that the perfect self-knower does not need to revise when her believed reliability differs from her (believed) degree of belief. The other would be to say that the subject with perfect knowledge of what her belief is, and perfect confidence about what her reliability on that matter is, should believe that the two are the same. That is, we could stipulate that a subject

---

$P(q \,.\, B) = P(q)\,P(B) = n\,(1 - \varepsilon)$, yielding

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

But where do we get that q and B are independent? In the argument above, this independence was guaranteed by the assumptions 2 and 3, that is, perfect confidence about both what one's degree of belief about q is and what one's reliability on that degree of belief is, because these together imply $P(B) = 1$.

$P(B) = 1$ implies that *every* proposition is independent of B, simply because of the way the extreme probability values work. A fortiori B is independent of q. This independence is thus the product of two factors, that we take it as a rational requirement that the subject have perfect confidence about what his degrees of belief are, and that we represent perfect confidence using probability.

with perfect knowledge of what her beliefs are should treat herself as already calibrated. The latter is what Bayesians who have made assumptions of perfect self-knowledge actually do think.

Re-calibration is not distorting

There are broadly three kinds of objection to re-calibration, that it makes one incoherent, that it is distorting, and that it is otiose, and there are various combinations of the three. I addressed one version of the charge of incoherence above, and will come back to the issue below. One of the first objections to Re-Cal that one hears is that it is redundant; there is no further information in the fact that you have degree of belief x in q than there was in the evidence for q that made you come to that degree of belief. Thus it should not change the degree of belief in q. It is a fact, though, that when we take information *about* our degrees of belief into account in the way set down in Re-Cal we often come up with a different first-order degree of belief in q. Thus, doing the re-calibrating procedure does not in general leave everything as it was, which suggests that the information at the second-order is distinct after all. One might protest further that this does not show there is new information at the second order. Rather, in re-calibrating one is double-counting the same first-order evidence, counting again evidence that one already counted, or should have counted, when one conditionalized at the first-order level. This is illegitimate because double-counting is illegitimate, and there should be no surprise that it often changes your first-order degree of belief in q. This explains both why re-calibration can change your degree of belief, and why it has no right to.

This sequence of objections depends on the persistent claim that the belief that you have a certain degree of belief in q contains no more or different information than does the evidence you used to get to that degree of belief in q. However, this is evidently false, because q and the proposition that you believe q have different contents. Accordingly, the evidence for them is different. q may be the proposition that dolphins don't smile, which is a claim about dolphins. By contrast, that you believe q to degree x is a claim about you, and not about dolphins. To investigate the first, you might read about dolphins. To investigate the second, you might think about how much you would bet *that* you have a certain degree of belief about dolphins. Neither kind of investigation would be suitable for the other proposition. That you believe q to degree x is a different proposition than q, and the evidence you have for them is different. Thus, there is no double-counting involved in taking both into account. Re-Cal can in general change one's degree of belief in q, but this is not via double-counting. It is because one is taking further evidence into account.

It is instructive to draw out the fact that there is an assumption that would make the claims q and S believes q to degree x susceptible to the same evidence, and so have the same content in the sense pertinent to the double-counting issue. This is the assumption that S believes q to degree q if and only if S fully believes that she believes q to degree x. If so, then whatever evidence is making her believe q to a degree is also relevant and sufficient for establishing *that* she believes q to that degree. In that case to count the fact that she believes q to some degree as an additional datum clearly is to re-count the same information. However, the assumption we used to get to this conclusion is a strong one that crosses first- and second- orders. It is also recognizable as the claim of perfect knowledge of one's own beliefs that we left behind in developing this framework in the first place. Between-order perfect self-knowledge claims are highly consequential for our topics here, and it is very easy to miss that one is making them.

Another common objection to re-calibration is that one can imagine a situation in which one has very little information about one's reliability on q, so, one might think, to change one's belief in q – one that can be imagined to be based on very strong evidence – on the basis of this weak evidence couldn't be right. For example, one might have only a track record of q-like judgments and that track record contains only one case. If one believes q to degree .7 and there was only one time in the past that one believed q to degree .7, and in that case q was true, then because one's .7 degrees of belief were associated with q being true in 100% of such cases, Re-Cal appears to counsel one to change one's degree of belief in q to 100%. Such a large change in degree of belief in q via a reliability judgment that is based on one case surely can't be right. Re-calibration is distorting.

The idea here is right, but Re-Cal also accommodates it. We must distinguish between what claim the evidence supports and how well it supports it. The 100% track record comes in as a piece of evidence for the value 1 for y in the reliability judgment $PR(q/P_i(q) = x) = y$. But that y equals 1 in that formula does not imply that one has 100% confidence in the claim "$PR(q/P_i(q) = 1)$." The confidence one has in that reliability claim will depend on how good that evidence that said y = 1 was. It wasn't very good in the case imagined, and Bayesianism will handle that here as it does in all other cases of evidence of varying quality. What one will typically have is a probability density function of confidences in various values for y, and consequently a probability density function for one's resulting confidence in q.[14]

Another factor that determines how much or little a single piece of evidence is worth is the subject's prior confidences about her reliability, and how well those are supported. Re-Cal is a rule of conditionalization in a Bayesian framework, and however Bayesianism handles the potency of a single data point to change one's posterior in a given situation is how it is handled here. How one's confidence in the fairness of a coin changes in response to a single coin toss, for

[14] This means that application of Re-Cal will usually be a Jeffrey conditionalization, but this is not surprising or problematic. In that Jeffrey conditionalization that weak reliability judgment will have less ability to affect the final degree of belief in q, just as we would want. We already saw above that the use of Re-Cal will also often be Jeffrey due to the fact that one natural way of failing to have perfect self-knowledge is to fail to have perfect confidence about what you degrees of belief, and one of these perfect properties, Confidence, Accuracy, or perfect confidence about one's level of reliability, has to fail if one is to use Re-Cal and remain coherent against the argument the argument discussed above.

example, will be different depending on the variance of one's existing probability density – with higher variance a bigger effect – and that is just as it should be both at the first-order and at the second.

Small data sets can in the short run, of course, lead one away from the true value of one's reliability and thus, via Re-Cal, away from the true probability of q on day i. Only in the long run could their bias be guaranteed to wash out (under further favorable supplies of evidence). However, this is no surprise given the nature of inductive reasoning, and is displayed just as prominently in subjective Bayesianism at the first order as it is in our second-order rule. Inductive reasoning is non-monotonic, defeasible, erodable. It is legitimate and common for one's confidence in q to go up and down with each step at which new evidence is accumulated. An inductive inference is legitimate only relative to a given body of evidence, and the erodability of such an inference continues no matter how much evidence one accumulates and how confident one becomes, except at the point where one's evidence deductively implies the conclusion. For the rationality issue on which the Bayesian can instruct the subject what to do, the question is what is the proper degree of belief to have in q given everything one knows now, not what it would be in the infinite long run when one would have all the evidence one does not now possess. Small or weak data sets will be handled differently in different situations depending on one's priors and their quality, but it is not difficult for a Bayesian framework to handle them. One cannot reject this 2^{nd}-order rule on the grounds of distortion by small data sets unless one also rejects the 1^{st} order Bayesian conditionalization rule.

The fact that we are dealing with a conditionalization rule neutralizes another intuitively appealing objection (Seidenfeld). This objection is that one could never do the Re-Cal step because one does not *know* one's calibration curve, that is, the real relation between one's reliability on q and one's confidence in q. Moreover, if one *did* know this curve one could calculate the true probability of q from it and one's actual confidence in q. Re-Cal is a nice rule if one can apply it, but if one had the information to do so, one would not *need* the rule! However, this complaint has a counterpart at the first order, which is that we do not know the true values of all of the elements we must estimate to apply first-order conditionalization. And it is not a complaint a Bayesian can afford to be swayed by, for requiring that we know the true values of all the things we need in order to conditionalize at the first order would imply requiring that we already know the true probability of the hypothesis we are using evidence to learn about. We are not required to know our calibration curve to use Re-Cal, but only to have evidence *relevant* to our reliability.

Seidenfeld (1985, 278) has objected that calibration is not what makes a forecaster's broadcasts informative to us. Weatherman A will be perfectly calibrated by announcing a 20% chance of rain day after day if 20% is the overall percentage of days per year that it rains in his location. Because this yields no discrimination as to which days are which, this would tell us very little on any *given* day about whether it will rain. By contrast Weatherman B may be quite reliable – when he says 99% the probability of rain is high -- but also uncalibrated – the probability of rain in that case is not 99% but 80%. He is overconfident across the board, but because the overconfidence is uniform across all his confidence levels about rain his announcements discriminate between the days when it is more likely to rain and the days when it is less. It seems obvious that the latter weatherman is more informative.

These are clearly possibilities, but the comparison is not probative for the re-calibration under discussion here. The claim of this paper is not that calibration will wash away one's other epistemological sins, but at most that a re-calibrated subject is better off than an *otherwise equal* subject who did not re-calibrate. That weatherman B is reliable means that he has information or methods for judging whether it will rain on a given, specific day. He has more than the annual rainfall figures for his town. There is no reason to expect that being perfectly calibrated would make up for Weatherman A's lack of that information about whether it will rain.

The comparison that is appropriate is between two subjects of equal reliability and different calibration levels. In the formulation of the framework here, subjects A and B have the same reliability for a given level of confidence when:

$$PR(q/P_A(q) = x_1) = PR(q/P_B(q) = x_1)$$

q is exactly as likely to be true when A believes it to degree x as when B believes it to degree x. A and B will have to be equally reliable on another level of confidence as well in order to make the comparison of calibration levels. Thus:

$$PR(q/P_A(q) = x_2) = PR(q/P_B(q) = x_2)$$

Suppose the difference between them is that A is calibrated and B is not. Thus:

$$PR(q/P_A(q) = x_1) = x_1 \text{ and } PR(q/P_A(q) = x_2) = x_2$$

but

$$PR(q/P_B(q) = x_1) > x_1 \text{ and } PR(q/P_B(q) = x_2) > x_2$$

Put this way, it is clear what is happening here. Although the two subjects are equally reliable about the weather, the calibrated subject is a more valuable source of weather information. This is because his confidence in reporting whether it will rain is no greater than the information that he has would warrant. The overconfident weatherman presents to we who depend on his forecast a higher confidence than his evidence warrants. The first presents to us the information he has, the second presents as having more than he actually has. Calibration by itself cannot make a subject a source of information – for that he needs to have information – but whatever information he does have will be faithfully conveyed to an otherwise ignorant observer only if the subject is calibrated.[15] Calibration has a crucial role in making a weatherman informative to

[15] This is so, of course, only if the observer has no independent evidence about the testifier's reliability. As a matter of psychological fact, human beings assume calibration by default. (refs) We often must make some assumption or

us. If we want information from others, then we should want everyone to be calibrated so that they present as having neither more nor less information than they actually do have.

The persistent worry that re-calibration is distorting has another source in a mistaken impression that the subject is free to choose how he will maximize his calibration. One imagines that since the 20% annual rate of rain in the subject's location is more securely known than any particular distribution of particular days of rain over the year, he will and may choose to report 20% confidence in rain every day rather than any more discriminating predictions, as the surest way to be calibrated. In other words one can imagine that he is permitted to maximize calibration at the expense of informativeness by hedging his bets. This is the assumption behind the argument that calibration is not a proper scoring rule. (ref)

However, there is no such latitude in Re-Cal because this rule gives a unique answer to what your resulting confidence in rain on each day should be given a set of evidence about rain that day and about your degree of belief and reliability, and the Principle of Total Evidence specifies that you must take into account the set of evidence that you actually have. The weather forecaster is permitted to maximize calibration by believing to degree 20% every day only if he has no evidence at all about rain on a particular day or about his reliability than what can be gleaned from the annual statistic for his location. If indeed this is the only evidence he has, then there is nothing objectionable about reporting 20% confidence every day on which he is so handicapped. Indeed, that is the rational thing to do, and does not qualify as hedging bets. When he has further evidence than the annual statistic, this will come in the form of evidence about today's prospects for rain, evidence about his confidence about that, and evidence about his reliability. He will first-order conditionalize on the evidence about rain particular to the day and come to a confidence about rain. He will observe that he has that confidence and consult any reliability information he has about how often it rains when he has that particular confidence, and re-calibrate. The chances of a person who has such evidence and does this ending up betting 20% every day is next to nil. The person has information beyond the annual statistic and is obligated by Re-Cal to use it, yielding a unique confidence that he does not have a choice about. The person who bets 20% every day is either ignorant of any particular information beyond the general annual rain statistic – in which case his behavior is not hedging – or he has more information and is not using it – in which case he is hedging, but also violating the Principle of Total Evidence. In no case does Re-Cal permit the weatherman to hedge his bets in order to attain calibration because in no case is he permitted to ignore evidence.

One might worry that Re-Cal brings distortion in another way, namely by interfering with the convergence to truth that we know you would have had in the long run had you done only

other because we have no information to scrutinize it in cases where we don't know the testifier or his track record. This default assumption is defeasible (refs), but the fact that we make it makes the actual calibration level of the testifier even more crucial to whether we are getting information from him.

first-order conditionalization. A convergence theorem in which we assume that both first-order and second-order evidence of the sort I propose are taken into account suffices to address this worry. Such a theorem can be proved by the likelihood-ratio method introduced by James Hawthorne (refs), a version of the Weak Law of Large Numbers, in which we can put a lower bound on the rate of convergence by putting a weak lower bound on the quality of information that is being supplied by the evidence sequences coming in, under the usual assumptions such as that prior probabilities are not extreme, pieces of evidence are independent given the truth of the hypothesis, and so on.[16] The quality of that evidence is measured by the log of the likelihood ratio, the ratio of the probability of that evidence coming up given the hypothesis and given the negation of the hypothesis. Applying this method in our current case would show that to the extent that the evidence about our own beliefs and about our reliability is of high quality, it also leads to convergence to the truth about q, that is to the objective probability of q. [Safe to skip this section and start again on p. 29 if not interested in proof details.]

The key to showing that this method will yield the usual result even for our new kind of evidence is to show that it is plausible for belief and reliability information to be of a sufficiently high quality, in the sense intended, to confirm $PR(q) = r$ over $PR(q) \neq r$. This sense is captured in the following expression:

$$\log[P(e^n/PR(q) = r.b.c^n)/P(e^n/PR(q) \neq r.b.c^n)] \qquad \text{QI}$$

The argument in this log term is the probability of e^n given that $PR(q) = r$ divided by that given that $PR(q) \neq r$, where b is background assumptions. e^n is a sequence of length n, a stream of evidence coming from a stream of observations or experiments c^n of length n, both of which are initial segments of an infinite stream. In our case, each entry, e_n, of e^n takes the form:

$$P(q) = x_i \,.\, PR(q/P(q) = x_i) = y$$

If QI is positive then the sequentially assimilated outcomes of the first n observations or experiments make the likelihood ratio for e^n and $PR(q) = r$ greater than that for e^n and $PR(q) \neq r$, thus discriminating the two hypotheses. QI is higher the higher that likelihood ratio is. This sequence, and so the value of n, is growing with the addition of each new observation. The theorem says that if QI is above a certain value, and its average expected value and variance have cooperative values, then the rate of convergence – roughly value of n by which all false alternative hypotheses are disconfirmed – has a specifiable lower bound.

[16] The beauty of Hawthorne's style of convergence theorem lies also in its lack of need for some assumptions that are typical for other convergence theorems; for example, this theorem does not require countable additivity, or that evidence consist of identically distributed events, and the explicit lower bounds on convergence means that there is no need to wait for the infinite long run. This convergence result applies even when agents make non-Bayesian (evidence-independent) revisions of the prior probabilities of hypotheses. This convergence theorem says that provided that such reassessments do not continually drive the prior probability of the true hypothesis ever closer to 0, the posterior probabilities of each of the true hypothesis's false competitors must approach 0 as evidence increases. (ref)

The question now is why we should expect the QI for our type of evidence to be greater than 0. Under what realistic conditions would q's objective probability being r rather than not r make some sequence of belief + reliability conjunctions more likely to be what the evidence stream turns up? For a simple case of this, suppose that q is not an indexed hypothesis like "rain today" but an eternal hypothesis. Then it has a true probability, call it r. Suppose that *PR* is frequency probability and we are estimating it via track record evidence. Then, although the subject may show varying degrees of belief x_i on each observation, if the true probability of q is r, then *PR*(q/P(q) = x), the frequency of the subject finding q when he goes to look whether q happened when he believed it to degree x_i will approximate r. As the track record piles up, the r trend should come to dominate. That is, if *PR*(q) = r, then QI for q and our type of evidence is likely to be greater than 0 because y is likelier to be equal to r than it is to be any of the other possible values.

If there are no such streams or they give out at some point then this kind of evidence isn't going to make the subject who uses it converge to the appropriate probability for q, but then no Bayesian should expect it to; all convergence theorems depend on the assumption that the subject will have evidence that has particular nice properties and comes in forever, and the only question that remains is whether those particular properties are realistic. That is, the success in the long run of the second-order conditionalization I propose at getting the subject to the true probability of q depends on the quality of his evidence. This is no surprise and is equally true at the first-order.

Why should we expect that there's a y-stream which is more likely if q is true rather than false? An easy way to see why is to imagine PR to be a frequency probability and to imagine y to be the true probability of q (which, assume for the moment, does not change with each passing day ans it would if q were "rain today"). Then the question is why if these pairs are coming in we should expect the PR terms to come to approximate the true probability. This is easy, provided the background and circumstances aren't deceptive in some way: PR(q/P(q) = x) is measuring how often q happens when a certain indicator – namely a particular belief-state of the subject – is present. *The true probability of q will determine how often q happens when the subject believes to that degree*, and since that the subject believes to that degree is assumed, the conditional probability

PR(q/P(q) = x)

is discharged to yield:

PR(q) = y

We cannot expect that on every occasion that the subject has that degree of belief that y will show up in the reliability term. Rather a sequence of outcome pairs that suggests y can be expected if y is the true probability of q and nothing deceptive is going on.

[PROOF DONE]

The convergence theorem shows that following Re-Cal (in addition to first-order conditionalization) will not lead you away from the true probability of q in the long run. However, there are convergence theorems showing that first-order conditionalization alone will get you to the true probability of q in the long run. Why is the extra labor of any value? Isn't re-calibration *otiose*? Sure, it looks more rational to get back in line with the Principal Principle when one appears to be out, but is there any less pious and more concrete benefit?

Sketch: There are several. One is that Re-Cal allows one to revise extreme degrees of belief, which it often seems there should be a way to do. Note that in:

$P_f(q) = P_i(q/P_i(q) = x \,.\, PR(q/P_i(q) = x) = y) = y$

X may be 1, while y is not 1. I.e., just because you are certain of q does not mean the objective probability of q is 1, or that your evidence gives grounds to believe you are perfectly reliable about q. This is easiest to understand with empirical propositions. You may be certain of q, but on reflection realize that you've not always been right when you were certain in the past. (We might be able to do something along these lines with necessary truths too, but it depends on whether it is coherent to believe that $PR(q) < 1$ for q a logical truth. It is clear that it is not coherent to have degree of belief in q that is less than 1, but this expression is different, so maybe).

Another added value of re-calibration is shown in the fact that you will get to the truth in the long run by using only Re-Cal and foregoing 1st-order conditionalization. It is completely unrealistic to think we would get enough higher-order evidence to make all of those conditionalizations possible, but this fact does further illustrate that re-calibration is distinct from updating at the first-order.

The concrete value of the fact that Re-Cal can move us along without first-order conditionalization is in the fact that there can be (short-run) situations where you don't have first-order evidence but you do have second-order evidence. The tiger described earlier provides an obvious case of this since the subject had no new information about tigers, but had information at the second order that could save her life. Another such situation is one where a person or community is making an assumption and treating it as unfalsifiable, but is not aware of doing so or of which assumption it is treating in this way. Via Re-Cal you will be able to correct for the effect of the false assumptions on your predictions, without even needing to identify what the false assumption is.