

“I’m onto something!”

Learning about the world by learning what I think about it

Maria Lasonen-Aarnio

***Analytic Philosophy*, Volume 56, Issue 4, winner of the
Mark Sanders Prize in Philosophy of Mind**

Abstract

There has been a lot of discussion about whether a subject has a special sort of access to her own mental states, different in important ways from her access to the states of others. But assuming that subjects can genuinely find out about their own minds, is the kind of *import* of acquiring self-knowledge different in some interesting, principled way from the import of finding out about the mental states of others? Consider, in particular, the import of finding out about the doxastic states of others who share your evidence. Recently many have argued that evidence about the opinions of others can provide both evidence about one’s evidence, and evidence about first-order matters that the evidence bears on. So, for instance, learning that a friend who shares my evidence is very confident that p can give me evidence that my evidence supports p , and evidence that p is true. But assuming that my own states are not perfectly luminous to me, could learning what I think about a matter have the same kind of evidential import? For instance, could learning that I am confident that p give me more evidence about whether p ? It is very tempting to think that evidence about my own doxastic states is inert in a way that evidence about the states of others is not. I argue that this is wrong: there is no principled difference between the evidential import of these two kinds of evidence. Asking what I think about a matter can be a perfectly legitimate way of gaining more evidence about it.

I Finding out about your evidence

Consider evidence about evidence, evidence that bears on the question of what some body of evidence supports. Few would deny that if I simply don’t know what my friend’s evidence is, but I do know that she is very good at evaluating such evidence and that she is very confident in p , this gives me evidence that her body of evidence, whatever it is, supports p . And it may also give me evidence that p . But more interesting are cases in which I have (and know that I have) the very body of evidence on which my friend’s opinion is based: in such cases, does the evidence I have screen out the evidential import that learning about her opinion might otherwise have?

The idea that such screening out always takes place has not been a very popular view of late. Consider somewhat standard situations of *peer disagreement*. Assume that you and your friend have evaluated a common body of evidence concerning the outcome of the next US presidential election. You evaluate the evidence correctly, becoming confident in d , the proposition that a Democrat will

win the race. You then learn that your friend is confident that a Republican will win. Many would argue that this gives you some higher-order evidence that your original evidence didn't support *d*, and that it would be rational for you to respond by diminishing your confidence in *d*.¹

You hear Madame Babineaux make a pronouncement. If your French is a bit rusty, or you are sitting far away, learning what others came to believe based on her testimony can give you insight into exactly what it is she said. Similarly, our access to our own evidence is often limited. At least sometimes learning what others make of the evidence can provide a further epistemic window into its testimony. For the purposes of this paper I will assume that this is right, and even applies to subjects who have evaluated their evidence perfectly correctly. I take the basic insight to be not merely that we are fallible evaluators of evidence, but that evidence need not have perfect, or perhaps even decent, access to its own testimony. Even if a body of evidence makes it likely that Hilary Clinton will win the next presidential race, it need not be certain, or perhaps even likely, on the evidence that Clinton is likely to win on the evidence. After all, if it was always rational to be certain just what the testimony of one's evidence was, it is not clear how information concerning the opinions of others could give a rational subject higher-order evidence about her evidence.

But now assume, further, that my own opinions are not perfectly luminous to me in the following sense: I do not always know what opinions I hold, and it is not always rational for me to be certain that I hold the opinions I do. Even subjects who are infallible about their own opinions might fail to satisfy such luminosity, so long as they don't know, or it is not rational for them to be certain, of their own infallibility.² Hardly anyone these days would defend perfect luminosity for actual subjects. Some still think that there is an interesting notion of rationality that requires such perfect luminosity.³ I disagree, but won't need to settle the issue. I will simply assume that the mere fact that we lack perfect luminosity does not put us so beyond the pale of rationality as to make it uninteresting to ask how we ought to respond to evidence about our own doxastic states.

So here is the question I want to ask: if I can learn about my *own* opinions, can such learning have the sort of evidential import that, many have argued, learning about the opinions of others can have? For instance, can learning that I am very confident that a Democrat will win the next presidential race give me evidence

¹ This has been a popular view in the literature on peer disagreement. See, for instance Elga (2007), Christensen (2007), Kelly (2005, 2010). The reader will notice that I am assuming there to be objective facts about evidential support. However, nothing I say essentially relies on there always being a uniquely rational opinion that reflects these facts – at least not as long as we can make sense of the idea that learning of the opinions of others can give evidence about which opinions are rational without assuming such uniqueness.

² Cf. Christensen (2007). (By “infallible” I mean that if a subject is certain that she holds a given opinion, she does in fact do so.)

³ Others assume merely that rationality requires having *good*, though not perfect, access to one's own credences: a rational subject's estimates of her own credences cannot be too far off (for instance, Egan & Elga 2005). Note that such an assumption is enough to raise the question I will be interested in.

that the evidence on which my confidence is based strongly supports a Democrat winning? Moreover, can it give me evidence that a Democrat will, indeed, win? Many feel an immediate urge to answer these question – especially the latter one – with an emphatic “no”: at least setting aside cases involving some abnormal causal relationship between how things are out in the world and my mental states, evidence about my present opinions is evidentially inert in a way that evidence about the opinions of others is not.⁴ It just seems absurd that I could ask myself what I think, and boost my confidence in a claim further as a result of learning about my own opinion. In this spirit, David Christensen (2011), for instance, writes:

Consider first how an agent should regard the information that she herself has reached a certain conclusion from her evidence. Suppose I do some calculations in my head and become reasonably confident of the answer 43. I then reflect on the fact that I just got 43. It does not seem that this reflection should occasion any change in my confidence. On the other hand, suppose I learn that my reliable friend got 43. This, it seems, should make me more confident in my answer. Similarly, if I learn that my friend got 45, this should make me less confident.

Now, it may be that Christensen’s remark is premised on an assumption of luminosity: I cannot genuinely learn that I reached a certain conclusion, since upon reaching that conclusion I automatically come to be certain that I did so. But let us proceed on the assumption that luminosity sometimes fails. In a similar spirit, Tom Kelly (2005) writes:

But notice that, when *you* enumerate the reasons why *you* believe that H is true, you will list the various first-order considerations that speak in favor of H—but presumably, *not* the fact that you yourself believe that H is true. From your perspective, the fact that you believe as you do is the *result* of your assessment of the probative force of the first-order evidence: it is not one more piece of evidence to be placed alongside the rest.

Kelly isn’t, I take it, merely making a point about what you *would* list as your evidence for the relevant hypothesis, but about what your evidence for the hypothesis in fact consists in. This, Kelly argues, raises a prima facie puzzle for so-called conciliationist views of disagreement: why should learning what you think about a question be evidence bearing on that question for *me*, if it is not such evidence for *you*?

I will dub the idea that evidence about my present opinions is inert in a way in which evidence about the opinions of others is not the *asymmetry view*. Note that those who hold the asymmetry view may be willing to concede that the opinions of my *past* self (assuming that those opinions are suitably independent of my present ones) can sometimes function like the opinions of other subjects. And maybe, if I have completely lost a body of evidence (perhaps due to forgetting) and all I now know is what I believe based on it, my belief can give me evidence about evidence I no longer have. But, the thought goes, when I *have* a body of evidence – at least

⁴ Roush (2009), however, is an exception.

setting aside abnormal cases – any such epistemic import of evidence about my own states is screened out. If Madame Babineaux is shouting in my ear, asking what *I* make of her speech cannot provide me with further clues as to what she is saying!

This initial reaction may seem to be further confirmed by reflecting on some of the odd-seeming consequences of allowing information about my own doxastic states to have the same kind of import as evidence about the states of others is taken by many to have. Wouldn't boosting my confidence in a proposition as a result of learning what I think involve some illegitimate recycling of my evidence? Couldn't I keep repeating the procedure, becoming more and more rationally confident in a proposition? And wouldn't one end up with violations of a plausible synchronic version of the Principle of Reflection?

Implausible as denying the asymmetry view may seem, it is not clear why the kind of reasoning outlined above should not apply in the first-person case. I consider my peers to be good evaluators of evidence. That is why information about their opinions gives me higher-order evidence about the import of our shared evidence. But I also consider myself to be a good evaluator of evidence. Why, then, cannot learning about my own opinions give me further evidence about the import of my own evidence? In what follows I will argue that this simple argument is essentially correct: it is at best implausible and at worst incoherent to allow evidence about the (present) doxastic states of others to have a certain kind of epistemic import, but not to allow evidence about one's own (present) states to have that kind of import. Though I will focus on doxastic states, my guess is that the kind of symmetry I will argue for is much more general. Assume that learning that a friend feels moral revulsion towards a certain action can give me evidence that it is wrong; that learning that she fears the glacier we are about to cross can give me evidence that it is dangerous; and even that learning that she enjoys *A Love Supreme* can give me evidence that it is an excellent record. Then, the kinds of arguments I will give can be expected to generalise to these other cases as well. So, for instance, learning that *I* feel revulsion toward an action could give me evidence that it is morally wrong.

Here is how I will proceed. I will first formulate two theses, *First-order inertness* and *Higher-order inertness*, that are attempts to capture the ideas that evidence about my own opinion about whether *p* is inert when it comes to the first-order question of whether *p*, as well as higher-order questions about what my evidence supports. As it turns out, there are arguments against these theses, many of which rely on assumptions that numerous epistemologists have recently accepted. After laying out my initial case, I consider a refined asymmetry view. Roughly, the idea will be that if I am rational, and have no reason to think that I am prone to either over- or under-estimate the import of my evidence, then at least *First-order inertness* will hold. I argue that even this refined version of the asymmetry view is false. Before concluding, I consider and answer some remaining qualms and objections.

II The inertness theses

As indicated above, I am assuming that learning about the opinions of others can have both *higher-order* and *first-order* import. For instance, if I discover that my peer believes that a Democrat will win the presidential race, this can give me both evidence about what our common body of evidence supports, and evidence that a Democrat will win the race. What I want to argue is that if this is right, then the same is true of learning of one's own present opinions. First-person evidence is neither higher-order nor first-order inert.

Let me begin by clarifying what I mean by talk of learning about someone's present opinions. Assume that it is now time t , and my peer and I hold opinions about whether d , opinions that are based on evaluating a common body of evidence. At a slightly later time t' we disclose what opinions we held at t . This in itself gives us new evidence, and in so far as the new evidence is not inert, it may no longer be rational for us to hold the opinions we did. Hence, if we are rational, the opinions we disclose might no longer be opinions we hold at t' , once we have taken the new evidence about the opinions we held at t into account. Similarly, assume that at a time t I am wondering how confident I am that d . Learning about my present opinion will be learning what my opinion is at t . But of course, by the time the learning takes place (at a later time t'), t will no longer be the present moment.

Evidence that is inert regarding a certain issue is evidence that doesn't make a difference for that issue. As far as one's opinions concerning the issue go, it is rational not to be swayed in any way by such evidence. When updating happens by conditionalization on new evidence, such inertness can be captured in terms of probabilistic independence. So, for instance, if a piece of evidence E' is inert as far as the question of whether d is true goes, d is probabilistically independent of E' : the probability of d conditional on E' just is the unconditional probability of d (and vice versa). Drawing on this thought, a first pass at the ideas that evidence about one's own states is both higher-order and first-order inert can be formulated as follows. Let Cre_t be the credence function of a (rational) subject at a time t . As indicated above, I am assuming there to be objective facts about evidential support. However, nothing I say essentially relies on there always being a uniquely rational way of assigning credences given a body of evidence. Let P_o be the (or a) rational credence function, a credence function that is appropriate given one's evidence at t . I will follow the convention of using lower-case letters for non-rigid designators: " $\text{cre}_t(p)=r$ " should be read as "my credence in p at t is r ", and " $\text{p}_o(p)=r$ " as "the objective, rational credence in p at t is r " (or perhaps, as "one of the rationally permitted credences in p at t is r "). Here, then, are the first-pass inertness theses:

First-order inertness

$$\text{Cre}_t(p|\text{cre}_t(p)=r)=\text{Cre}_t(p)$$

Higher-order inertness

$$\text{Cre}_t(\text{p}_o(p)=r'|\text{cre}_t(p)=r)=\text{Cre}_t(\text{p}_o(p)=r')$$

I will assume that the proponent of inertness will allow plugging in logically weaker proposition about what my credence in p at t is. For instance, it would not be in the spirit of the asymmetry view to allow p to be probabilistically dependent on the proposition that I have a *high* credence in p , or that my credence is between 0.4 and 0.6.

Unfortunately, the above inertness theses run into immediate problems. Some qualifications will be needed before we have a view that does not falter at the very outset. Consider *First-order inertness*, and let c be as follows:

c : I assign a credence of 0.8 to some proposition at t .⁵

Assume that, not having perfect access to my credences, at t my credence in c is 0.9, but I am not certain of this – for all I know my credence in c might, instead, be 0.8. But conditional on my credence in c being exactly 0.8 at t , my credence in c ought to be 1. Hence, $\text{Cre}_t(c)=0.9$, but $\text{Cre}_t(c|\text{cre}_t(c)=0.8)=1$. Hence, if we let c be *any* proposition whatsoever, it isn't too difficult to generate counterexamples to *First-order inertness*.⁶ A restriction must be placed on the propositions that the thesis is to apply to.

But even restricting the propositions that the inertness theses are to apply to won't quite do it. Assume that I am told by highly reputable sources that neuroscientists have ensured, possibly by manipulating my credences, that the following is true: my credence in a proposition is high just in case the proposition is true, and low just in case the proposition is false. Even those who defend the asymmetry view might well admit that such cases are exceptions, for I have evidence that there is an abnormal causal dependence between whether p and my credence in p . The same kind of point can be made to apply to *Higher-order inertness*: assume that I am told by highly reputable sources that neuroscientists have ensured that my credence in p is 0.9 just in case my evidence supports p to degree 0.5. Then, if I have enough reason to trust the testimony, wouldn't learning that I am 0.9 confident in p give me evidence that my evidence only supports p to degree 0.5?

A further need for refinement might arise from noting that *First-order inertness* is incompatible with a version of van Fraassen's *Principle of Reflection* that applies to one's current credences. Yet, such a principle has seemed to many to be

⁵ In so far as there are uncountably many precise credences I could have, it may be that I ought to assign a credence of 0 to the proposition that I assign a credence of *exactly* 0.8 to some proposition. To avoid such issues, we can assume that talk of assigning a credence of 0.8 to some proposition should be understood as assigning a credence that is within some non-zero interval containing values both below and above 0.8.

⁶ Aaron Bronfman (2009) makes a structurally similar point in a slightly different context. Note also that not all of the problematic propositions are strictly about my own credences: a similar problem arises if we let p be the proposition that *some* subject assigns a credence of 0.8 to some proposition (assuming it is not already rational for me to be certain of this proposition).

an uncontroversial part of *Reflection*⁷, capable of avoiding the kinds of counterexamples that a more general version faces:

Current Reflection

$$\text{Cre}_t(p|\text{cre}_t(p)=r)=r$$

Note that such a principle faces the same kinds of counterexamples as an unrestricted version of *First-order inertness*.⁸ Hence it, too, would have to be restricted. But to see that the two principles are incompatible, assume, for instance, that $\text{Cre}_t(p)=0.8$. If the subject doesn't know what her credence is, then for some $r \neq 0.8$, she will assign a non-zero credence to $\text{Cre}_t(p)=r$. Then, the subject cannot satisfy both *Current Reflection* and *First-person inertness*: if $\text{Cre}_t(p|\text{cre}_t(p)=r)=r$, she will satisfy the former but violate the latter; if $\text{Cre}_t(p|\text{cre}_t(p)=r)=0.8$, she will satisfy the latter but violate the former. Nevertheless, someone might be drawn to both *Current Reflection* and the idea of inertness for essentially the same reason, namely, that subjects should respect their own opinions in the way captured by *Current Reflection*, which entails that learning about one's own opinion should not occasion a change in that opinion.⁹ Indeed, perhaps the idea behind inertness all along was merely that *learning* propositions about one's doxastic states cannot have the relevant sorts of first- or higher-order import, where it is assumed that only true propositions can be learnt.

In light of this, friends of inertness might want to add the qualification that the above theses should only be taken to apply to those values r that one's credence in p in fact takes (assuming, again, that updating happens by conditionalization). If *First-order inertness* is restricted in this way, then satisfying it will never force violations of *Current Reflection*. In fact, *Current Reflection* entails *First-order inertness* thus restricted. Then, any counterexample to the latter will also be a counterexample to the former. Indeed, my arguments below, targeted at *First-order inertness*, will also be arguments against *Current Reflection*.

I will now argue that the inertness theses – even if restricted in the ways discussed above – are false. I first state three arguments the premises of which at least a lot of people are committed to. The first deploys the premise that it is sometimes possible for a rational agent to have evidence that (s)he is prone to over- or under-estimate the force of her evidence. The second uses probabilistic reasoning to show that if one regards oneself as a reliable evaluator of the evidence in a particular way, then conditionalizing on information about one's own credences is bound to lead to violations of *Higher-order inertness*. The third points out that popular views of peer disagreement commit one to the falsity of the inertness theses. These arguments should make the idea that evidence about one's own states

⁷ Van Fraassen (1984: 248) himself, for instance, remarks that the synchronic version of Reflection should be uncontroversial.

⁸ Let c be as above, and assume that I assign a non-zero credence to $\text{cre}_t(c)=0.8$. Again, conditional on my credence in c being exactly 0.8 at t , my credence in c ought to be 1. But then, we get a violation of *Current Reflection*, for $\text{Cre}_t(c|\text{cre}_t(c)=0.8) \neq 0.8$.

⁹ See Christensen's (2007) criticism of Dutch Book arguments in favour of Current Reflection.

is inert start looking a lot less plausible. After laying out these arguments, I consider a refined version of the asymmetry view.

III Against the inertness of evidence about one's own doxastic states

(i) Chandra's prediction

It is possible to have evidence that one's present evaluation of the truth of a particular proposition is too pessimistic (or too optimistic). Consider the following case:

Chandra's prediction 1

Chandra has spent his life predicting the outcomes of political elections. Based on a vast body of evidence E_{ORIGINAL} , he has formed a credence in the proposition that a Democrat will win the next presidential race (proposition d). An angel whom Chandra has every reason to trust tells him that he has a strong tendency to slightly under-estimate the prospects of Democratic candidates. Unfortunately – the angel adds – this tendency will tend to persist even once he learns about it. Chandra then learns that he is 80% confident that a Democrat will win.¹⁰

Consider time t' at which Chandra has received the testimony of the angel, but has not yet learnt about his own credence. We can assume that despite the trustworthiness of the angel, her testimony is misleading, and Chandra has a perfectly rational credence in d . Still, it would seem rational for Chandra to violate *First-order inertness* in the following way: his credence in d , conditional on his credence being 0.8, ought to be above 0.8.

Someone might object that disapproving of one's own current credences as too low, in the way that Chandra does, is incompatible with being perfectly rational. Here one might try to deploy arguments to the effect that it is irrational to regard oneself as an *anti-expert*, as someone whose beliefs are far from the truth (cf. Sorensen 1987, Egan & Elga 2005). So, for instance, Sorensen (1987) argues that it is irrational to believe that one is an anti-expert about p , while also being aware that

¹⁰ There are similar cases in the literature in which subjects seem to violate *First-order inertness*, even if these cases are not presented as counterexamples to the thesis. Take, for instance, Christensen's (2007) example of a subject who thinks she is overly optimistic about the weather. Letting s be the proposition that tomorrow will be sunny, $\text{Cre}(s|\text{cre}(s)=0.8)=0.6$. Christensen uses this case to argue for a principle of moderate self-respect (more moderate than *Current Reflection*), since the agent is probabilistically incoherent if 1) she violates *Current Reflection* in the way described, 2) she assigns a sufficiently high credence to $\text{cre}(s)=0.8$ (in Christensen's example, she assigns a credence of 0.9 to this proposition), and 3) it is in fact the case that $\text{cre}(s) = 0.8$. However, only these three claims together imply that the agent is probabilistically incoherent, and not 1) alone. Hence, Christensen's (2007) argument gives us no reason to think that violating *First-order inertness* in itself makes a subject irrational.

one believes p .¹¹ But to argue that Chandra is guilty of self-ascribing anti-expertise, we must decide what it would be for credences, as opposed to all-out beliefs, to display anti-expertise. A natural suggestion is that one's credence in p displays anti-expertise when it is sufficiently inaccurate: for instance, if one is confident in p , when p is false.¹² And a natural way of cashing out what it is to self-ascribe such anti-expertise is in terms of conditional probabilities: one has a low credence in p , conditional on having a high credence in p (for instance, $\text{Cre}(p|\text{cre}(p)>0.8)<0.2$). If, in addition, one is sufficiently confident of having a high credence in p , one is certainly incoherent. But it was not assumed that Chandra regards his credence in d as displaying anti-expertise in this sense. In effect, we can fill out the details of *Chandra's prediction 1* so as to make his credences perfectly coherent.^{13 14}

To say the least, it is far from clear how arguments against the rationality of self-ascribing anti-expertise could be deployed to show that Chandra is irrational. Besides, at least to show that *Higher-order inertness* is incorrect, we can, instead, consider a case in which a subject acquires evidence that his credences are, at least within a limited domain, perfectly rational:

Chandra's prediction 2

Chandra has spent his life predicting the outcomes of political elections. Based on a vast body of evidence E_{ORIGINAL} , he has formed a credence in the proposition that a Democrat will win the presidential race (proposition d). An angel whom Chandra has every reason to trust tells Chandra that his credence in d is (and will continue to be) perfectly rational. At this point

¹¹ Sorensen (1987) distinguishes between *commissive* and *ommissive* anti-experts about a proposition p . In the former case, one's belief that p is strong evidence that $\sim p$, and one's belief that $\sim p$ is strong evidence that p . In the latter, p is true just in case one does not believe it.

¹² Egan & Elga (2005) spell out anti-expertise by means of the notion of inaccuracy, though they are concerned with being an anti-expert about a subject matter – thought of as a list of propositions – rather than a particular proposition. An anti-expert about a subject matter has sufficiently inaccurate credences in sufficiently many propositions in the list. For instance, one might think that a subject is an anti-expert about subject matter if she has high confidence in at least some of the propositions representing the subject matter, and at least half of these propositions are false. But note that *Chandra's prediction 1* need not involve any such self-ascription of anti-expertise.

¹³ Letting d be the proposition that a Democrat will win, and Cre be Chandra's credence function at time t' just before he learns of his own 0.8 confidence in d , we can, for instance, assume that $\text{Cre}(d|\text{cre}(d)=0.8)=0.875$. There are numerous credence distributions Chandra might have concerning what his own credence in d is that would make him probabilistically coherent. Here is one: Chandra is 0.9 confident that his credence in d is 0.8 (the credence he actually has), and 0.1 confident that his credence is 0.114. Moreover, because he takes himself to systematically under-estimate, $\text{Cre}(d|\text{cre}(d)=0.114)=0.125$. Then, $\text{Cre}(d)=\text{Cre}(d|\text{cre}(d)=0.8) \times \text{Cre}(\text{cre}(d)=0.8) + \text{Cre}(d|\text{cre}(d)=0.114) \times \text{Cre}(\text{cre}(d)=0.114) = 0.875 \times 0.9 + 0.125 \times 0.1 = 0.8$.

¹⁴ Roush (2009) discusses this sort of problem as a problem for *Current Reflection*, suggesting a restriction of the principle that would make it immune to the counterexample. The idea is, roughly, that counterexamples involve background evidence, like evidence that one tends to over- or under-estimate. However, even if we apply the Roush-style restriction to the inertness theses, this would not support the asymmetry view. After all, when evidence about the doxastic states of our friends is not inert, we seem to always have some background evidence, such as evidence about their reliability – indeed, Roush (2009: 260) argues for this very conclusion.

Chandra is neither sure what his own credence is, nor what the rational credence (or range of rational credences) in d is. He then learns that he is 80% confident in d .

Assume that the angel is in fact right: Chandra's credence in d (both initially, and after the testimony of the angel) is perfectly rational. Assuming that Chandra has very strong reason to trust the angel, conditional on his credence in d being some value r , he is virtually certain that r is a rational credence in d . However, before learning about his own confidence he is not at all certain that a confidence of 0.8 in d is rational. In this case it seems that learning about his own credence could *at least* give Chandra evidence that it is rational to assign a credence of 0.8 to d based on his original evidence.

Finally, consider a more mundane case in which Chandra merely regards his doxastic states as reliably, though not infallibly, reflecting his evidence. Perhaps he knows how reliable he is through the testimony of the angel, perhaps he considers his own track record, or perhaps his views about his own reliability have some other rational basis. If learning about his own doxastic states can at least have higher-order evidential import when he knows, or is certain, that his credences are correct, it would seem bizarre if they couldn't do so when Chandra merely thinks that it is likely that his credences are correct.¹⁵ In the next section I give an argument making this idea more precise. I argue that if one conditionalizes on evidence about one's credences, and regards oneself as above 50% likely to assign the rational credence to b , whatever that credence is, then at least *Higher-order inertness* has got to be false.

(ii) Regarding oneself as a reliable evaluator of the evidence

Given a certain way of cashing out what it would be to take oneself to be a reliable evaluator of one's evidence, conditionalizing on information about one's own credences is at least sometimes bound to have higher-order import.

Assume that having evaluated a body of evidence E_{ORIGINAL} , I am uncertain what the import of my evidence is regarding a proposition p . That is, I am uncertain what credence it is rational for me to assign to p in my current situation. I have a credence distribution over a partition of hypotheses $\{O_1, \dots, O_n\}$ about what the rational credence in p is given that evidence (2. below). There is no need to think of each hypothesis as stating, of a point-valued credence, that it is the rational credence in p . Rather, given that the set of hypotheses is finite, we can think of each as stating that the rational credence in p lies within a certain range. So, for instance, one of the hypotheses might state that the rational credence in p lies within the

¹⁵ We might also consider an intermediate case: if the angel tells Chandra that he has a perfectly ideal credence, but Chandra thinks there is a slight (10%) chance that the angel is lying, then upon learning, for instance, that his credence a Democrat will win is 0.95, shouldn't Chandra be roughly 90% confident that this credence is ideal? How does this differ from a case in which Chandra merely regards his own credences as 90% reliable, and then learns of his 0.95 credence?

interval $[0, 0.1[$. Or, it might state simply that the rational credence in p is low.¹⁶ Similarly, $\{M_1, \dots, M_n\}$ form a partition of hypotheses about my credence in p (assumption 3. below). Though the formal result does not depend on how a hypothesis O_i and a hypothesis M_i are related, assume that these hypotheses are coordinated. For instance, if O_i is the hypothesis that the rational credence in p is 0.9, then M_i is the hypothesis that my credence in p is 0.9. If O_i is the hypothesis that the rational credence lies within a given the interval, M_i is the hypothesis that my credence lies within that interval. Slightly more formally, assume that there are n disjoint non-empty sets of values in the interval $[0, 1]$, and that these sets can be completely ordered. For instance, set 1 might consist of all the values in the interval $[0, 0.1[$, set 2 of all the values in the interval $[0.1, 0.2[$, etc. Each hypothesis O_i states that the rational credence in p is a member of set i . Each hypotheses M_i states that my credence in p is a member of set i .

The third assumption will be that the probability of M_i conditional on O_i is over 0.5. In light of the above coordination assumption, this amounts to a reliability assumption about how my own credences track the rational credences: no matter which hypothesis about what it is rational to believe is correct, I am likely to assign to p a credence that is in line with the hypothesis. So, for instance, conditional on a hypothesis stating that the rational credence in p is within the interval $[0, 0.1[$, I am at least 50% likely to assign to p a credence that is within this interval. Or, conditional on a hypothesis stating that it is rational to be confident in p , I am at least 50% likely to be confident in p . It would seem perfectly rational for a subject to take her credences to track the rational credences in this manner.

With 'V' for disjunction, the assumptions made are the following, where Cre is my credence function after evaluating my evidence, but prior to learning what my own credence in the relevant proposition p is:

1. $Cre(M_i|O_i) > 0.5$
2. $Cre(\bigvee_{1 \leq i \leq n} (O_i)) = 1$, for all $i \neq j$, $Cre(O_i \& O_j) = 0$, and for all i , $0 < Cre(O_i)$
3. $Cre(\bigvee_{1 \leq i \leq n} (M_i)) = 1$, for all $i \neq j$, $Cre(M_i \& M_j) = 0$, and for all i , $0 < Cre(M_i)$

But 1. and 2. entail that for arbitrary $i \in \{1, \dots, n\}$

1. $Cre(O_i|M_i) > Cre(O_i)$.¹⁷

Hence, conditionalizing on any hypothesis M_i concerning the value that my own credence in p takes leads to boosting my credence in O_i , a hypotheses about the rational credence in p . Hence, conditionalizing on information about my own

¹⁶ A certain degree of permissibility can be accommodated by taking the hypotheses in question to state that the permissible credences are exactly those that lie within a given range. We will, however, need to assume that the ranges specified by the different hypotheses don't overlap.

¹⁷ The proof is in Appendix 1.

credences at least has higher-order import and hence, my credences will fail to comply with *Higher-order inertness*.

It certainly seems possible for the above assumptions to hold for a rational subject. Consider, for example, the following simple case. I am not sure to what extent my evidence supports the proposition p , but I do know that either it makes p likely, or it makes p unlikely (perhaps this much has been revealed to me by an epistemology oracle). Hence, the partition of hypotheses about the rational credence only contains two members. I am not sure what I myself think about the matter, but I treat my credences as tracking my evidence with over 50% success: conditional on my evidence making p likely, I regard myself as over 50% likely to have a high degree of confidence in p , and conditional on my evidence making p unlikely, I regard myself as over 50% likely to have a low degree of confidence in p . It follows that merely learning that I am confident (unconfident) in p should make me boost my confidence that my evidence supports p to a high (low) degree.

To conclude my tentative case against the inertness theses, I will say why popular views of peer disagreement seem to entail that evidence about one's own opinions is not inert.

(iii) Disagreement

One of the nefarious activities of Madame Babineaux is exploiting the linguistic ineptitude of vegetarian tourists by making them order bone marrow soup in her Parisian bistro. My friends and I did our best to place our order, and to explain what a vegan is, and the Madame just repeated, amid the very loud chatter, what dish she has us all down for. Let E_{ORIGINAL} be our common body of evidence about what dish it is that she is about to send our way, and b the proposition that we are about to receive some of the infamous soup. At a time t_0 I have done my best to determine whether or not b is true, but unfortunately, I have no better access to my own opinion than I do to the opinions that others might hold about the issue.¹⁸ In fact, the evidence makes b likely (say to degree 0.9).

Assume that at a slightly later time t_1 , I learn that whereas one of my friends, Pro-peer is confident that b , I am confident that $\neg b$. Hence, my total body of evidence now consists of

E₁:

- E_{ORIGINAL}
- Pro-peer is confident that b , whereas I am confident that $\neg b$.

However, if the inertness theses hold, then as far as the relevant propositions go (propositions about how likely b is on my evidence, and b itself), whether or not I learn what *my* opinion is should not make any difference. Consider, then, a counterfactual situation in which instead of learning both what Pro-peer thinks and

¹⁸ I take this to be compatible with the thought that I may have different kinds of evidence about my own opinions and the opinions of others; it's just that I don't have *better* evidence about my own opinion.

what I think, I simply learn what Pro-peer thinks and hence, a counterfactual situation in which my total evidence consists of the following:

E_1^* :

- E_{ORIGINAL}
- Pro-peer is confident that b .

If evidence about my own opinions is inert, then as far as the relevant propositions go, it doesn't make a difference whether my total evidence at t_1 is E_1 or E_1^* .

But now compare bodies of evidence E_1^* and E_1 . The former consists of the original evidence E_{ORIGINAL} , which supports p , and of the evidence that Pro-peer is confident that b . Given natural assumptions, E_1^* does not support p to a lesser degree than E_1 – after all, that Pro-peer is confident that b is, if anything, evidence that the original evidence supports b , which it in fact does. By contrast, consider E_1 . The evidential situation I am in when I have evidence E_1 is a somewhat standard peer disagreement case. According to many who hold so-called conciliatory views, I should give at least some weight to my own opinion – indeed, Elga's (2007) Equal Weight View, for instance, urges giving my own opinion and that of my peer equal weights, and what White (2009) dubs the "thermometer model", I should treat my own credences as a guide to what is true in exactly the same way that I treat your credences. Many others have also argued that I ought to assign at least some weight to both opinions, ending up with a confidence in b that is somewhere between the two.¹⁹ By contrast, it seems that evidence E_1 makes it rational to be at least as confident in p as Pro-peer is. Hence, given standard conciliatorist views of peer disagreement, evidence E_1 and evidence E_1^* make reasonable different attitudes toward b , as well as toward propositions about which credence in b is rational. It follows that evidence about my opinion cannot be inert. Hence, the thesis of the inertness of first-person evidence appears to be incompatible with popular views about peer disagreement.²⁰

It is worth also noting a somewhat bizarre consequence of the inertness view. First, consider a case in which I know, prior to gaining information about anyone's opinions, that I am Con-peer. I then, at time t_1 , learn that whereas Pro-peer is confident that b , Con-peer is confident that $\neg b$. Hence, my total evidence now consists of

¹⁹ Note that Christensen (2011) defends the idea that evidence about my own states is inert. This entails that E_1 and E_1^* support the relevant proposition (in this case b) to the same degree, and that E_1^* doesn't, it seems, support b to a lower degree than E_{ORIGINAL} alone. Hence, he endorses the conclusion that the rational attitude to b based on E_1 is no lower than that based on just E_{ORIGINAL} .

²⁰ In the example given it was assumed that my original opinion was in fact irrational. As Christensen (2011: 4) points out, conciliatorists are not committed to saying that if a subject takes evidence about peer disagreement into account as they recommend, she automatically ends up with a rational opinion. However, even if we take conciliatorism to be a view about how a particular kind of evidence (i.e. evidence about peer disagreement) ought to be taken into account, the point made shows conciliatorism to be incompatible with inertness of first-person evidence.

E₁:**

- E_{ORIGINAL}
- Pro-peer is confident that b , whereas Con-peer is confident that $\neg b$.

Since I know that I am Con-peer, by inertness, learning about Con-peer's opinion should have no effect on my credences in the relevant propositions. As far as these propositions go, it is as if my total evidence was just **E₁***. Now consider a counterfactual situation in which I first (at t_1) learn about the opinions of Pro-peer and Con-peer, and *then* learn (at a yet later time t_2) that I am Con-peer. Assuming evidence to be commutative, the order in which I learn these two items shouldn't matter. Again, it is as if my total evidence in the end was just **E₁***. But then, I ought to discount Con-peer's opinion merely as a result of learning that I am Con-peer. This strikes me as bizarre: why should merely learning that I am Con-peer suddenly render Con-peer's opinion inert? What justifies refusing to give a judgment any weight at all just because it is my own?²¹

Having presented a tentative case against the inertness theses, I will now discuss a way of trying to re-formulate an asymmetry between the evidential import of third- and first-person doxastic states that acknowledges the arguments given above. This will also allow me to complete my argument against *First-order inertness*.

IV A refined version of first-order inertness

Perhaps the arguments given above still leave room for an interesting asymmetry between the epistemic import of evidence about one's own doxastic states and evidence about the doxastic states of others. Consider the following case:

The boost

Based on evaluating a body of evidence E_{ORIGINAL}, I become fairly confident that a Democrat will win the next Presidential race (proposition d). I have no reason to think that I am prone to either over- or under-estimate the force of my evidence. Upon learning of my own credence in d , I further boost my confidence, becoming very confident in d .

There is something particularly implausible about the idea that such a boost in my confidence could be rational. Perhaps denying that this could happen is, in the end, at the heart of the idea that evidence about one's own doxastic states is inert in a way that evidence about the states of others is not: even if I am perfectly rational and have no reason to think that you tend to over- or under-estimate the force of

²¹ The issue can perhaps be sharpened by assuming that I learn (and hence, become certain) that Con-peer is a molecule for molecule duplicate of me and hence, that Con-peer and I have exactly the same credences at t_0 . Then, if the information that I am confident that $\neg b$ is inert, information that Con-peer is confident that $\neg b$ must also be inert (since I am certain that Con-peer is confident that $\neg b$ just in case I am confident that $\neg b$). Similarly, in coming to learn that Con-peer is my duplicate, I should completely discount his opinions. But why should I discount Con-peer's opinion just because she is a duplicate of myself? Thanks to Billy Dunaway for bringing up the issue of duplication.

your evidence, then learning what you think about whether p can still make it rational for me to change my opinion regarding p . But if I am perfectly rational, and have no reason to think that I am prone to over- or under-estimate, then learning what I think about p cannot make it rational for me to change my opinion. It may seem that nothing I have said so far would constitute an argument against this refined and qualified version of the asymmetry view.²²

Note that the refined view concedes that evidence about my own opinions can have the same kind of import as evidence about the opinions of others. That is, at least in some cases it can bear on both first-order and higher-order questions. As such, the view concedes much of what I set out to argue for. Still, I think that even the refined asymmetry view is false. For unless some of the arguments given above are faulted, it must be conceded that evidence about one's own (rational) opinion can still bear on *higher-order* questions about what one's evidence supports. And then, the viability of the proposed view requires being able to pull apart first- and higher-order evidential import: upon learning of my credence in d , it might be rational for me to change my confidence in various higher-order hypotheses about what the rational confidence in d is, but it is not rational for him to change his confidence in d .

First, it is simply false that evidence bearing on higher-order questions about the degree to which the evidence supports various propositions never has any bearing on first-order propositions. In general, evidence that one's evidence supports a proposition p is evidence for p . Here is a somewhat heuristic way to see why. That it will rain tomorrow raises the probability that you will have evidence supporting rain, and vice versa. But then, we should expect evidence bearing on whether the evidence supports rain to bear, in rather typical cases, on whether it will rain today. It is true that probabilistic relevance is not in general transitive, but circumstances have to conspire in a rather special way to block such transitivity. Besides, to refute the claim that the first- and higher-order levels are completely insulated, it is enough to show that there is *some* possible evidence that bears both on a first-order question (such as the question of whether a Democrat will win the next Presidential race), as well as a higher-order question (such as the question of how likely it is on one's evidence that a Democrat will win).

Hence, the proponent of the refined view had better not rely on a general thesis stating that the first- and higher-order levels are insulated. There is, however, a train of thought worth exploring, an argument for the view that learning about my own rational credences can have higher-order import, but it cannot have first-order import a view. It relies on the assumption that a rational subject's credence in a proposition p will match her *expectation* of what credence it is rational for her to have in p . Assume that I have a perfectly rational credence of 0.9 in a proposition p . On the refined inertness view, if I regard myself as reliable, then learning that I am 0.9 confident in p may be evidence that 0.9 is the ideal, rational credence in p given my evidence. I should take some credence away from hypotheses stating that the rational credence is something other than 0.9, and move all that credence to the 0.9

²² For instance, in the peer-disagreement case I considered above, it was assumed that the opinion I held to start out with was irrational.

hypothesis. But assume, first, that (i) my initial credence of 0.9 in p was my expectation of the correct, ideal credence. And assume, second, that when I learn that my own credence is 0.9, (ii) I take credence away from other hypotheses in such a way that my expectation of the rational credence is 0.9 remains the same. Assume, in particular, that before learning of my own credence, my credence distribution over hypotheses about the rational credence could be represented by a Gaussian curve peaking on the 0.9 hypothesis. Upon learning of my own credence, I end with a Gaussian curve peaking on the 0.9 hypothesis, but it is now sharper, as I am more confident that that hypothesis is true. By the above assumptions, if my expectation of what the ideally rational credence in p is doesn't change, then whatever evidence I acquired cannot have relevance for whether p . The same kind of argument could be applied to cases involving evidence about the doxastic states of others: if I start out with an ideal credence of 0.9 in a proposition p , then learning that your credence is 0.9 shouldn't effect my credence in p , as long as I have no reason to think that you tend to either over- or under-estimate the import of the evidence.

One of the problems with the above argument is its reliance on a *rational reflection* principle stating that a rational subject's credence in a proposition p equals her expectation of the rational credence.²³ It is worth discussing a simple case where such a principle fails. What is interesting about the case is that it demonstrates that sometimes there is a very strong dependence between the first- and higher-order levels, so that *any* evidence bearing on a first-order proposition p also bears on some higher-order proposition about the degree to which the evidence supports p , and vice versa. Let propositions p and q be *strongly dependent* if any learning that changes the probability of one also changes the probability of the other. Assuming that updating happens by conditionalization, p and q are strongly dependent just in case $0 < \text{Cre}(p) < 1$, and for any e

$$\text{Cre}(p|e) \neq \text{Cre}(p) \text{ iff } \text{Cre}(q|e) \neq \text{Cre}(q).^{24}$$

There are what strike me as very convincing cases of strong dependence between first-order propositions and higher-order propositions. Such cases arise when just what one's evidence is (and hence, what credence distribution it is rational to have) depends on how things stand in the world. Consider, for instance, *Clock Beliefs*.²⁵

You are looking at the minute hand of an unmarked clock from some distance away. The hand moves in discreet one-minute jumps. Given your perceptual abilities and your distance from the clock, you are not an infallible judge as to the exact position of the hand. Assume that the hand in fact points to 20 past the hour. If the

²³ See Christensen (2010) for a discussion of rational reflection principles. The principle Christensen dubs "Rational Reflection" entails the formulation I have given in terms of expectations, but is not entailed by it.

²⁴ Hence, a simple example of strongly dependent propositions are p and $\neg p$, when one is certain neither that p is true nor that it is false.

²⁵ Adapted from David Christensen's (2010) example, which is adapted from Williamson (2014). See also Elga (2013) and Lasonen-Aarnio (2015) for a discussion of such cases.

hand were to point to 19 or 21 past the hour, you would have a visual experience that was slightly different, but such differences are so small that you are not able to reliably tell which exact experience you are having. Then, given the nature of your perceptual evidence, it seems that it would not be rational for you to be certain that the minute hand points to 20 past the hour. For any relevant i , let p_i be the proposition that the minute hand points to i minutes past the hour. Assume, for instance, that in addition to the proposition that the hand points to 20 past the hour (p_{20}), you should assign some credence to both p_{19} and p_{21} .²⁶ Assume also that given your distance from the clock, your abilities of perceptual discrimination, etc., something similar is true in other cases: you always have a 1-minute “margin for error” in either direction (and we can assume that you know this). So, for instance, conditional on the hand pointing to 19 past the hour, it is rational for you to assign some credence to its pointing anywhere between 18 and 20 past the hour.

In the kind of example described, which credence it is rational to assign to propositions about the position of the minute hand depends on what the position of the minute hand in fact is. But not only this, there are pairs of first- and higher-order propositions that are *strongly dependent* in the sense introduced above: both are dependent on some propositions, and they are dependent on exactly the same propositions.²⁷ For any relevant i , let P_i be the credence distribution that is rational just in case p_i is true. Then, for any i , the (first-order) proposition p_i and the higher-order proposition that P_i is rational are strongly dependent. The reason why this is interesting in the present context is that it shows that there are cases in which any failure of *Higher-order inertness* will *inevitably* lead to a failure of *First-order inertness*: in such cases changing one’s confidence in higher-order propositions about which credence in some proposition p is rational will make it rationally mandatory to change one’s confidence in p .²⁸

In effect, a case of clock beliefs in which a subject is uncertain both about her own credences and about her own rationality, nevertheless taking her credences to track the ideally rational credences in a pretty weak way constitutes a counterexample to even the refined inertness thesis. In Appendix 2 I describe a model in which the following assumptions hold. First, it is uncertain whether one is rational: conditional on any proposition p_i (that is, conditional on the minute hand in fact pointing to i minutes past the hour), it is not certain that one’s credence function is P_i . Nevertheless, one’s credence function is never too far off from what is rational: conditional on p_i , one’s credence distribution is either P_{i-1} , P_i , or P_{i+1} . So, for

²⁶ Everything I say below is compatible with thinking that you ought to assign a higher credence to the case you are in fact in. So, for instance, if the clock points to 20 past the hour, perhaps your credences should be as follows: $\text{Cre}(p_{19})=0.25$, $\text{Cre}(p_{20})=0.5$, and $\text{Cre}(p_{21})=0.25$. All that I will need to assume, is, first, that $\text{Cre}(p_{19})=\text{Cre}(p_{21})$, and second, that $\text{Cre}(p_{19})\leq \text{Cre}(p_{20})$. So long as one concedes that it is not rational for a subject to be certain, of the case that she is in, that she is in it, these assumptions seem overwhelmingly plausible.

²⁷ Strong dependence entails dependence, but not vice versa, since (probabilistic) dependence is not a transitive relation: if p and q are dependent, and p is dependent on e , it does not follow that q is dependent on e .

²⁸ This is not to say that such strong dependence is necessary for failures of *Higher-order inertness* to lead to failures of *First-order inertness*; only that it is sufficient.

instance, conditional on the hand pointing to 20 past the hour, the subject's credence distribution is either P_{19} , P_{20} , or P_{21} . Second, conditional on P_i being the rational distribution, one's credence distribution is likelier to be P_i than $P_i(p_{i-1})$ or $P_i(p_{i+1})$. That is, one is likelier to have the rational credence distribution than either of the two possible irrational ones. This is fairly weak reliability assumption, weaker than the one I made in connection with the Bayesian argument given above, for the subject need not even be *likely* to have a rational credence distribution. As I show in Appendix 2, given these assumptions, conditionalizing on the information that one's own credence function is P_i raises the probability that p_i , as well as the probability that P_i is rational.

A credence function satisfying the assumptions just described would seem perfectly rational.²⁹ The subject doesn't have perfect discriminatory abilities, and she knows this. She also knows that what credences it is rational for her to have conditional on a hypothesis about the position of the minute hand depends on just what that hypothesis is. But then, taking herself to always have exactly the correct credence distribution would, it seems, be to assume that she can perfectly discriminate the case she is in from the adjacent cases after all. At the same time, the assumption that conditional on p_i , one's credence function is either P_i , P_{i-1} or P_{i+1} fits well with what was assumed about one's discriminatory abilities. Finally, it is easy to imagine background evidence – such as data about how reliably her own credences have tracked the ideal credences in clock belief type scenarios in the past – that would make it rational think that conditional on p_i , one is at least a bit likelier to have credence function P_i rather than P_{i-1} or P_{i+1} . Even the refined thesis of inertness fails, for if evidence about one's own doxastic states provides evidence about evidence, it sometimes also inevitably provides evidence about first-order matters such as the position of a minute hand of a clock.

Let me take stock. I stated some initial arguments against the inertness theses. Against these arguments, it looked like there might still be an interesting asymmetry between the epistemic import of evidence about the doxastic states of others' and evidence about one's own doxastic states. But the refined asymmetry view relied on being able to separate first- and higher-order epistemic import. I argued that such a separation cannot, in general, be effected: at least in some cases evidence that bears on first-order propositions also bears on higher-order propositions about what the evidence supports. I discussed *Clock Beliefs* as an example where there is a strong dependence between first-order propositions about the position of the minute hand of the clock and higher-order propositions about what credences it is rational to have. If one is uncertain about one's own rationality, but nevertheless takes one's credences to track the ideally rational

²⁹ It is worth noting that the example of clock beliefs described above is only coherent assuming uncertainty either about one's rationality or about one's own credences (or both). For it was assumed that the subject is certain, and rationally so, of the way in which the rational credences depend on the position of the minute hand. She knows, for instance, that P_{20} is rational just in case p_{20} is true. But then, if the subject is certain both that she is rational – and hence, that she assigns P_{20} just in case p_{20} – and of what her own credences are – and hence, that her own credence function is P_{20} – she must be certain of p_{20} . But it was assumed that it is *not* rational for the subject to be certain of the position of the minute hand.

credences in the weak way described, information about one's own credences is neither first- nor higher-order inert. Hence, even the refined asymmetry view fails.

Despite everything that has been said, one might still have various worries about a view on which evidence about one's own doxastic states has the kind of import I have argued for. I will now address some of these.

V Some remaining objections

(i) The objections from double counting and informational proxies

Recall the case *First-person boost* in which I increase my confidence in a proposition upon learning that I am fairly confident that it is true. I have heard several people react to such cases by saying that they involve some illegitimate double counting or recycling of one's evidence.³⁰ What goes on, according to this objection, is not too different from the kind of flaw I commit in the following case:

Asha and Mosi

My belief about the temperature outside is based on Asha testifying that it is 25°C. I know that Mosi repeats whatever Asha says about the weather. Nevertheless, upon hearing Mosi say that it is 25°C, I boost my confidence that the temperature outside is 25°C.

Strictly speaking, this case doesn't involve counting the very same evidence twice. Rather, I count as relevant a piece of evidence that I know to be thoroughly dependent on evidence that I have already taken into account. This is what gives rise to the charge that I am double counting: I am more or less recycling or counting twice whatever evidence is provided by Asha's testimony. And note that even if Mosi only repeated what Asha said 80% of the time, tossing a coin about what to say the remaining 20%, Mosi's testimony could not give me new evidence about the temperature outside, and the charge of double-counting could be mounted. Either way, the two testimonies are not suitably independent.

What is the analogy between the above cases and cases such as *The boost*? Assume that at least if I am rational, my doxastic states are always based on how things *seem* to me. As such, my credence merely encodes that seeming. If I give evidential weight to my credence, I am in effect counting the seeming twice. And even if my credences don't perfectly track my seemings, the situation is analogous to the case in which Mosi repeats what Asha says most of the time, but sometimes tosses a coin about what to say. To make what is going on more vivid, imagine that there is a little homunculus in my mind. Whenever I am in a new evidential situation, the homunculus makes a recommendation. Perhaps, for instance, the recommendations take the form of "0.9 that *p*!". When things go well, my credences reflect the recommendations of the homunculus. If I then adjust my opinion as a

³⁰ Kelly (2005) mentions the idea that even counting the doxastic responses of *others* as more evidence involves an illegitimate sort of double counting.

result of learning of my own credences, it is as if I was counting the recommendations made by the homunculus twice, and in so doing treating the opinion of a single informant as the opinions of two independent informants.

Presumably, on the kind of model just sketched we are to think of the seemings – i.e. the testimony of the homunculus – as evidence. The suggestion might be that ultimately, at any time all of our evidence consists of such seemings, or of the testimony of the homunculus.³¹ But if so, then the above line of thought could not be used to support the asymmetry view, for it would also show that taking into account the doxastic states of *others* who share one’s evidence involves an illegitimate kind of double counting. Assume that it seems to me that it is raining and hence, that my homunculus says “It is raining!”, or perhaps “0.9 that it is raining!”. If my friend and I have the same relevant evidence, then our homunculi are saying the same things. If I know that we have the same evidence, and that our credences just track (when things go well) the testimony of our respective homunculi, then by the above reasoning, in taking into account her credence I am engaged in exactly the same kind of double counting. Evidence about the opinions of others who share my evidence is also, it seems, rendered inert.

Perhaps even more importantly, if one’s evidence consists of seemings, then to assume that it is always clear what the import of the seemings are, or what one’s homunculus testifies to, is to assume that it is always clear what one’s evidence is. In the case described above, Mosi’s testimony doesn’t yield any evidence about the import of Asha’s testimony, or about exactly what it was she said. But even if Mosi repeats what Asha says, this need not be the case: if Asha shouts into the wind and is barely audible, Mosi’s testimony might well have evidential relevance. If there is uncertainty about the testimony of the homunculus, and my credences are very reliable at tracking it, then some further argument must be given for the conclusion that learning about those credences couldn’t provide an epistemic window into that testimony.³²

What has been said also defuses a related objection:

“Belief-states are a kind of informational proxy for one’s evidence, just as transcripts of documents are proxies of originals, or pictures of scenes are proxies of those scenes. The evidential force of such proxies of imperfect representations is screened out by the things they are representations of”.

Sometimes a transcript of a document misrepresents the original. And even when it doesn’t, any information gained by reading it can also be gained by reading the

³¹ Some epistemologists think that evidence consists of seemings. See, for instance Huemer’s (2007) *phenomenal conservatism*.

³² Would the homunculus-model look better if we didn’t think of the seemings in question as evidence? The idea would be that my only access to my own evidence is via the testimony of my homunculus, but that testimony is not itself part of the evidence. It is difficult to see what epistemic role the testimony plays on such a view. If I base my beliefs on the testimony, then how can they be justified, since the testimony is not part of my evidence? Moreover, if the testimony is not part of my evidence, it is not clear why information about the testimony would be inert – and my credences could certainly provide information about the testimony.

original. Either way, in the presence of the original document the transcript is evidentially inert – or so goes the objection.

However, it is not true that the evidential force of a proxy is always swamped by the thing it is a proxy for. For instance, assume that an original document is written by hand, and that it is difficult to make out the writing. The transcript, by contrast, is typed. You know that the person who typed it was reliable, though far from infallible, at deciphering hand-written documents. Hence, their transcript is only an incomplete representation of the original. Nevertheless, the transcript can provide evidence that is not screened off by the presence of the original document: your access to what the original document says is limited, and the transcript is a source of information about what the original says.

The objection from informational proxies has an even more obvious problem, namely, that it generalizes too far: if a doxastic state is a proxy that loses any evidential force it might have in the presence of the evidence it is a proxy for, then information about other subjects' beliefs would be evidentially inert.³³ But we were looking for a problem that is specific to one's own doxastic states.

(ii) The objection from single-sidedness

Learning about the opinions of others can provide a genuine check on one's own opinions: the rationality or correctness of one's opinions is confirmed when others agree, and disconfirmed when they disagree. One worry is that evidence about one's own doxastic states doesn't have this dual aspect, and for that reason, cannot function as genuine evidence³⁴:

“If I am confident in p , then learning of my own confidence is evidence that my evidence supports p , as well as evidence that p is true. But had I been confident in $\sim p$, learning about my own confidence would have provided me with evidence that my evidence supports $\sim p$, as well as evidence that p is false. So learning about my own doxastic states could only ever give me evidence confirming the correctness of those states, but never evidence disconfirming their correctness. But surely it is a *criterion* of proper evidence that it can confirm or disconfirm the correctness of an opinion.”

There appear to be several issues that this objection is running together.

Here is one way of construing the above worry. Assume that at a time t you are confident in p , and at a slightly later time t' you learn of your confidence in p . Now, it may be true that if you had no reason to think that you are either prone to over- or under-estimate, then the learning that takes place at t' cannot give you

³³ Kelly (2005) discusses an argument along these lines leading to the conclusion that even evidence about the doxastic states of *others* is inert.

³⁴ For instance, Christensen (2011: 7) says something along these lines: “we may take the first-person psychological evidence to be incapable of providing the sort of check on one's reasoning that third-person evidence provides. In this sense, it is relatively inert”. However, by no means am I attributing the objection I formulate to Christensen.

evidence that your state of confident in p at t was irrational. By contrast, learning about the opinions of others can do this. This is a trivial corollary of the fact that a subject can learn that the states of others differ from her own doxastic states. However, conceding this is different from accepting any sort of inertness thesis, for the issue is whether it might be rational to change one's confidence in various propositions (including p) as a result of the learning that takes place at t' . Perhaps learning of my own doxastic states cannot give me evidence that those states are irrational, but so what? It was never part of my claim that evidence about one's own doxastic states and evidence about the doxastic states of others has exactly the same sort of import.

Here is another way of construing the worry. It has often been proposed that if the outcome of a test or procedure is to be evidentially relevant for some hypothesis, then it must be capable of both confirming and disconfirming that hypothesis.³⁵ Perhaps, then, the worry is that as I have set things up, evidence about my own states is only capable of confirming certain hypotheses. However, it is far from clear whether this can be used to discredit the kind of symmetry view I am defending. The view I am defending does *not* entail anything like the general claim that for any proposition p , if I learn that I believe p (or, perhaps, that I am above 0.5 confident in p), then I can boost my confidence in p . Moreover, it is not clear what consequences the criterion on evidential relevance just stated has even against a view that *does* entail this. For there need not be any one hypothesis, whether about the degree to which my evidence supports a proposition p , about p itself, or about my own rationality, that is confirmed no matter what I learn about my doxastic states. A proposition p may be confirmed if I learn that I am confident in p , but it would have been disconfirmed had I been confident in $\sim p$. And nothing that has been said entails that I should become more confident that my credence in p (whatever it is) is rational as a result of learning what that credence is. It is worth noting that in the kind of simple model I give in Appendix 2, learning what my own credence in the relevant proposition is does not confirm the proposition that my credence (whatever it is) is rational.³⁶

(iii) The objection from bootstrapping

One worry about the view I have been defending is that if I can boost my confidence in a proposition as a result of learning how confident I am in it, then this leads to the absurd consequence that I can bootstrap myself to a higher and higher confidence

³⁵ This has been one reason to discredit “bootstrapping” – reasoning to the reliability of one's faculties from premises that one knows or justifiably believes based on employing those very faculties – on the grounds that bootstrapping can only ever confirm the reliability of my faculties. Jim Pryor has defended this sort of response. See White (2006) for discussion.

³⁶ Once I have opened my eyes and seen the minute hand of the clock my credence function is P_{20} . Then, my credence that my credences are rational is $P_{20}(\{w_{19}^{19}, w_{20}^{20}, w_{21}^{21}\})$. Once I learn that my own credence function is P_{20} , my credence that my credences are rational is $P_{20}(\{w_{19}^{19}, w_{20}^{20}, w_{21}^{21}\} | \{w_{19}^{20}, w_{20}^{20}, w_{21}^{20}\}) = P_{20}(\{w_{20}^{20}\} | \{w_{19}^{20}, w_{20}^{20}, w_{21}^{20}\})$. I leave it as an exercise to the reader to check that $P_{20}(\{w_{20}^{20}\} | \{w_{19}^{20}, w_{20}^{20}, w_{21}^{20}\}) = P_{20}(\{w_{19}^{19}, w_{20}^{20}, w_{21}^{21}\})$.

by repeating the same process.³⁷ Perhaps, through enough repetitions, I can work up to a confidence that is arbitrarily close to 1.

Assume that at a time t_1 I learn the proposition $\text{cre}_0(p) = r_0$, and that this proposition is positively evidentially relevant for p . The objection assumes that at t_1 the proposition $\text{cre}_1(p) = r_1$ must likewise be positively relevant for p . But whether or not evidence about one's doxastic states has first-order evidential import depends on specific features of one's epistemic situation, and the fact that my situation at one time has such features doesn't in any way guarantee that those features will continue to be present at subsequent times. It is no part of the view I am defending that evidence about one's own states *always* has either first- or higher-order import. It is not difficult to construct cases in which $\text{cre}_0(p) = r_0$ is relevant for p at t_0 , $\text{cre}_1(-)$ results from conditionalising $\text{cre}_0(-)$ on information about one's credence in p at t_0 , but in which $\text{cre}_1(p) = r_1$ is not relevant for p at t_1 . A simple case of this sort is one in which upon learning $\text{cre}_0(p) = r_0$ at time t_1 , one also becomes certain of $\text{cre}_1(p) = r_1$. In such cases, the two propositions about one's credences at times t_0 and t_1 aren't genuinely two different pieces of information, for learning one is in effect learning both. But even when one genuinely acquires two different pieces of information, there is nothing inevitable about such information being relevant for p on both occasions.

Second, if I $\text{cre}_0(p) = r_0$ and $\text{cre}_1(p) = r_1$ are genuinely two different pieces of information for me, it is not clear why their both being relevant for p would be so absurd after all. Note that even if evidence about my own doxastic states continued to be positively evidentially relevant at subsequent times, it doesn't follow that I could eventually reach a confidence in p that is arbitrarily close to 1, for it may be that at each step I could increase my confidence less and less, and my credence started approaching some value below 1. The case would be very much like one in which I use the same measuring device twice in a row, and boost my confidence slightly as a result of reading the second measurement. If I keep doing this, it is plausible that at each round I acquire less and less evidence, and at some point observing the outcome of a new measurements might simply become evidentially irrelevant.

There is no easy step from denying the asymmetry view to the conclusion that by repeatedly learning of my own confidence in some proposition, I could always "bootstrap" myself to virtual certainty that that proposition is true.

(iv) The objection from the Transparency Thesis

Assume (perhaps contrary to fact) that there is a strong correlation between rain and the BBC weather forecast predicting rain. Then, based on seeing that it is raining, it may be perfectly rational for me to increase my confidence that last night's BBC forecast predicted rain. But now assume that I came to know that it will rain based on watching the BBC forecast. Then, I cannot increase my confidence that the BBC forecasts rain based on my knowledge that it will rain. This would involve an obvious, illegitimate kind of epistemic circularity. And it might seem similarly

³⁷ Thanks to Brian Weatherson for pressing this objection.

irrational to increase my confidence that my evidence supports the proposition that it will rain today. The objection now is that if some form of the Transparency Thesis is true, cases like *The boost* involve a similar kind of epistemic circularity.

According to defenders of the Transparency Thesis, when I wonder what my mental state regarding a proposition p is, I attend to precisely the same considerations as those I would attend to when I wonder whether p .³⁸ A typical idea is that I infer that I believe that p from p itself. But then, it would be problematic to increase my confidence in p as a result of coming to know that I believe p – this would be like believing that it is raining based on the BBC forecast, and then increasing my confidence that the BBC predicted rain as a result. The problem with this construal of transparency in the present context is that it has no obvious application to one's knowledge of doxastic states like suspension of judgment, or to more fine-grained credal states. "It is raining; therefore my credence that it is raining is 0.93" looks like a terrible piece of reasoning. But in so far as the objection is that the Transparency Thesis raises some general problem for those who deny inertness, there had better be a way of formulating the thesis that also applies to knowledge of doxastic states other than outright belief or disbelief.

Here is an alternative thought. Knowledge about one's doxastic state regarding a proposition p is based on the very same *evidence* as one's belief in p itself. The worry, then, is that if I base my belief that I believe p on a body of evidence E , I cannot then employ my knowledge of my belief as a basis for further beliefs about E or p . Now, it is much less obvious that this would be circular. But perhaps the problem is that again, we are dealing with an illegitimate kind of double counting. For if I believe p based on E , and I also believe, say, that I am confident in p based on E , then if I boost my confidence in p on the basis that I am confident in p , then it looks like I am spotting myself two separate reasons to believe p , where I really only have one reason, E .

This objection is only as strong as the Transparency Thesis it is based on. Hence, it is worth seeing how the transparency account is supposed to work. Fernandez (2006) makes the following proposal. Assume that there is a reliable connection between my evidence supporting a proposition to some degree and me assigning to that proposition a credence reflecting that degree of support. The thought, then, is that if my evidence E in fact supports p to degree 0.9, believing, based on E , that my credence in p is 0.9 is a reliable way of forming beliefs about my own credences. But why exactly is this? E , let us assume, is a set of propositions about the external world. Surely there need not be any connection whatsoever between these propositions being true and my credences. Nevertheless, because there is a reliable connection between *having* E as evidence and believing p to a certain degree, and I couldn't base my belief on E unless I had E as evidence, this method of forming beliefs is reliable. However, it is only available to me: you cannot justifiably form beliefs about my beliefs just based on E .³⁹

I am not at all convinced that any form of the Transparency Thesis is true. The above account of first-person access strikes me as problematic for numerous

³⁸ The Transparency Thesis is normally attributed to Gareth Evans (1982).

³⁹ See Fernandez (2003: 362-363).

reasons. Assume that my credences track degrees of evidential support fairly reliably, but only 80% of the time. It is far from intuitive that such a correlation would be reliable enough to yield *knowledge* of one's own mental states by the above kind of reasoning. Still, does my self-knowledge really dwindle if I become less reliable at proportioning my doxastic states to my evidence? It is far from obvious that it does. Further, the kind of reasoning given above seems to generalise too far. Consider the fact that there is (trivially) a very strong correlation between having *E* as evidence and having *E* as evidence. Then, by the above reasoning, believing that my evidence consists of exactly *E* based on my total evidence *E* would be a *very* reliable belief-forming process, since I couldn't, again, base a belief on *E* if I didn't have *E* as evidence. However, sometimes subjects are in no position to know just what their evidence is. Indeed, such access can vary wildly despite the ultra-reliable, trivial correlation between having *E* as evidence and having *E* as evidence. What, then, do such correlations have to do with access to one's evidence?

Note further that even those who defend the Transparency Thesis don't think that it is the *only* way of finding out about one's own doxastic states. I could instead resort to the kinds of methods by which I come to know about the mental states of others, by observing my own behaviour – or a neuroscientist could simply tell me what I believe. In such cases the kind of circularity worry raised would have no bite. And my point would still hold: it is false that evidence about the doxastic states of others can have certain kinds of evidential import that evidence about one's own states *cannot* have. Hence, even the Transparency Thesis would not establish that evidence about one's own opinions is inert. At most, it would establish that it is inert a lot of the time, and much more often evidence about the opinions of others. But even this would depend on a viable version of Transparency.

Conclusions

I have argued that in so far as evidence about another subject's doxastic states can provide evidence about what one's evidence supports, and about first-order matters such as who will win the next Presidential election, sometime evidence about one's own states can also do so. Evidence about one's own states is neither higher-order nor first-order inert. Sometimes asking what I think about a matter is a perfectly legitimate way of gaining more knowledge about it.

Let me conclude with a couple of rather speculative remarks. If what I have said is correct, then we have discovered a new kind of epistemic pathway by which one could find out about the world by finding out about one's beliefs. This pathway that doesn't in any way rely on those beliefs having specific types of externalist content, but merely on the assumption that they are reliable indicators of what the evidence supports, and (hence) reliable indicators of what is true.

Does this make for a kind of a priori evidence about matters external to the mind? This depends on whether subjects need some sort of empirical evidence in order to be justified in regarding themselves as reliable evaluators of evidence. But assume that at least sometimes they do not: they are entitled, without any further evidence, to regard their beliefs as fairly accurately reflecting the evidence, and

perhaps as fairly good guides to the world. Then, merely learning about my own doxastic states could give me new evidence about the world, evidence that would be at least as a priori as that provided by introspection, or whatever process it is by which I learn about my own mind.⁴⁰

⁴⁰ I am grateful to Billy Dunaway, Andy Egan, Jim Joyce, Brian Hedden, Sophie Horowitz, David Manley, Daniel Morgan, Sherri Roush, Josh Schechter, Teruji Thomas, Jonathan Vogel, Brian Weatherson, and participants at a colloquium at MIT in May 2012, at the Harvard Workshop on Belief in September 2012, at the Midwest Epistemology Workshop in September 2012, at a colloquium at the University of Southern California in November 2013, the Jowett Society in Oxford in June 2014, and a seminar at the University of St Andrews in June 2014.

Appendix 1

$\{O_1, \dots, O_n\}$ form a partition of hypotheses about what credence in a proposition p is rational on my evidence at a time t (2. below). $\{M_1, \dots, M_n\}$ form a partition of hypotheses about what my credence in p is at t (3. below). For any M_i , M_i is true just in case one's credence is in the range specified by O_i . Since the question concerns my credences, I will write 'P' instead of 'Cre'. With 'V' for disjunction, the assumptions made are then the following:

1. $P(M_i|O_i) > 0.5$
2. $P(\bigvee_{1 \leq i \leq n} O_i) = 1$, for all $i \neq j$, $P(O_i \& O_j) = 0$, and for all i , $0 < P(O_i)$
3. $P(\bigvee_{1 \leq i \leq n} M_i) = 1$, for all $i \neq j$, $P(M_i \& M_j) = 0$, and for all i , $0 < P(M_i)$

What I want to show is that for arbitrary $i \in \{1, \dots, n\}$

4. $P(O_i|M_i) > P(O_i)$.

Assume that $i = 1$.

$$5. P(M_1) = P(M_1|O_1) \times P(O_1) + \sum_{j>1} P(M_1|O_j) \times P(O_j)$$

$$6. P(M_1) < P(M_1|O_1) \times P(O_1) + \sum_{j>1} P(M_1|O_1) \times P(O_j) \quad (1., 5.)^{41}$$

$$7. P(M_1|O_1) \times P(O_1) + \sum_{j>1} P(M_1|O_1) \times P(O_j) = P(M_1|O_1) \sum_{j=1}^n P(O_j) = P(M_1|O_1) \quad (2.)$$

$$8. P(M_1) < P(M_1|O_1) \quad (6., 7.)$$

$$9. \frac{P(M_1|O_1)}{P(M_1)} > 1 \quad (8.)$$

$$10. P(O_1|M_1) = \frac{P(M_1|O_1)}{P(M_1)} \times P(O_1) > P(O_1) \quad (9.)$$

Hence, 3. is true.⁴²

⁴¹ It follows from 1.-3. that for $j \neq 1$ $P(M_1|O_j) < 0.5 < P(M_1|O_1)$

Appendix 2

It was argued that in a case of *Clock Beliefs* it is not rational for a subject to be certain of the position of the minute hand. Again, let p_i be the proposition that the minute hand points to i minutes past the hour, and let P_i be the credence function that is rational in a case in which one is looking at the minute hand of the clock, and it points to i minutes past the hour (that is, p_i is true). It was assumed that if the hand in fact points to i minutes past the hour, it is not rational for the subject to be certain that it doesn't point to $i-1$ or $i+1$ minutes past the hour.⁴³ Let us also assume that it is rational to assign an equal credence of $1/3$ to all three possibilities. This will, in effect, be required by the formal framework, which assumes updating to happen by conditionalizing a prior credence function on evidence thought of as a set of worlds, though it is worth noting that the proof I give below goes through even if this assumption is relaxed. (For the purposes of the proof, we can replace 2. below by $P_i(p_{i+1}) = P_i(p_{i-1})$ and $P_i(p_i) \geq P_i(p_{i+1})$).⁴⁴ Assume, in addition, that though one is uncertain that one's credences are rational, those credences track the ideally rational credences in the way already described in the main text (assumptions 3.-5. below).

The assumptions made about the case of *Clock Beliefs* described in which there is uncertainty about the position of the minute hand, about one's own credences, and about one's own rationality, are as follows. For any relevant i and j such that $|i - j| \leq 1$:

1. $P_i(p_{i-1} \text{ or } p_i \text{ or } p_{i+1}) = 1$
2. $P_i(p_i) = P_i(p_{i-1}) = P_i(p_{i+1})$
3. $P_i(\text{cre}=P_{j-1} \text{ or } \text{cre}=P_j \text{ or } \text{cre}=P_{j+1} | p_j) = 1$
4. $P_i(\text{cre}=P_{j-1} | p_j) = P_i(\text{cre}=P_{j+1} | p_j)$
5. $P_i(\text{cre}=P_j | p_j) > P_i(\text{cre}=P_{j-1} | p_j)$ and $P_i(\text{cre}=P_j | p_j) > P_i(\text{cre}=P_{j+1} | p_j)$

"cre= P_i " should be read as "one's credence function at t is P_i ", where t is the time at which one has opened one's eyes and is looking at the minute hand of the clock, but has not yet learnt what one's own credences are. In particular, no matter which case one is in at t , the above assumptions hold for one's credence function, as long as it is rational. That is why I write " P_i " instead of " Cre_i ". The actual case, we are assuming, is one in which p_{20} is true and hence, one's actual credence function is P_{20} .

⁴² I am grateful to Teruji Thomas for making this proof much simpler than it would otherwise have been.

⁴³ I assumed that there are only three (and not, for instance, five) hypotheses about the position of the minute hand to which it is rational to assign some non-zero credence merely for the sake of simplicity.

⁴⁴ That is, all that needs to be assumed for the proof to go through is, first, that the hypotheses that is true should not receive less credence than the others and second, that if, for instance, the minute hand points to 20 past the hour, equal credence ought to be assigned to the hypotheses that it points to 19 past the hour as the hypothesis that it points to 20 past the hour.

The above assumptions entail that conditionalizing on the information that one's credence function is P_i leads to increasing one's confidence that p_i and hence, in P_i . That is,

$$P_i(p_i | \text{cre} = P_i) > P_i(p_i)$$

For instance, conditionalising on the information that your own credence function is P_{20} should make you more confident that the minute hand points to 20 past the hour. Because in the model described the proposition p_{20} just is the proposition that the ideally rational credence function is P_{20} , it should also make you more confident that the rational credence function is P_{20} .

I want to show something even more general. In particular, replace 2. above by the following assumption (2*. entails 2.):

$$2*. \quad P_i(p_{i+1}) = P_i(p_{i-1}) \text{ and } P_i(p_i) \geq P_i(p_{i+1})$$

First,

$$\begin{aligned} \text{Let } P_i(p_i) &= r_1, \text{ and } P_i(p_{i+1}) = P_i(p_{i-1}) = r_2 \\ \text{Let } P_i(\text{cre} = P_{j-1} | p_j) &= P_i(\text{cre} = P_{j+1} | p_j) = 1/m \\ \text{Let } P_i(\text{cre} = P_j | p_j) &= 1/n \end{aligned}$$

By the above assumptions (1., 2*., 3., 4., and 5.),

$$\begin{aligned} m > n, \text{ and } 1/n + 2(1/m) &= 1 \\ r_1 &\geq 1/3 \\ r_1 + 2r_2 &= 1 \end{aligned}$$

Now,

$$P_i(p_i | \text{cre} = P_i) = \frac{r_1 \times 1/n}{r_1 \times 1/n + 2(r_2 \times 1/m)} > r_1$$

Hence,

$$P_i(p_i | \text{cre} = P_i) > P_i(p_i)$$

I will now describe a model (or, more precisely, a frame) exemplifying the case of *Clock Beliefs* described. The approach will draw heavily on that in Williamson (2011). A probabilistic frame $\langle W, R, P_{\text{PRIOR}} \rangle$ will consist of a set W , a relation R between members of W , and a probability distribution P_{PRIOR} over subsets of W . Informally, W is a set of situations or cases that we can think of as centered worlds. R is a relation of epistemic accessibility: wRw' just in case the subject in w cannot

epistemically rule out being in w' . I will assume that w' is epistemically accessible from w just in case being in w' is compatible with the subject's evidence at w . At least in the finite case, a rational subject assigns some non-zero credence to a case just in case it is epistemically accessible for her. Informally, we can think of P_{PRIOR} as one's probability distribution just before seeing the minute hand of the clock. I won't assume the distribution to be uniform, assigning to each member of W the same weight; however, I will assume that it is indifferent as to what the actual position of the minute hand is. To get the probability function P_w at a world w , we conditionalise P_{PRIOR} on the set of worlds accessible from w . Informally, we can think of this set as the total evidence one has at w , upon opening one's eyes and seeing the minute hand of the clock. Hence, the framework so far described assumes updating to happen by Bayesian conditionalization on a prior probability function.

Here is the frame $\langle W, R, P_{\text{PRIOR}} \rangle$ I will be interested in. Let " w_i^j " denote a world at which p_i is true (and hence, the minute hand points to i minutes past the hour), but at which the subject's credence function, upon opening her eyes and seeing the minute hand, is P_j (that is, the credence function that is rational just in case the hand points to j minutes past the hour). W consists of a set of worlds w_i^j such that $|i - j| \leq 1$ (for all relevant i). That is, every world in the model is such that if a proposition p_i is true and hence, credence function P_i is rational, the subject's credence function is either P_i , P_{i-1} , or P_{i+1} . For any i , P_i is gotten by conditionalizing the prior function P_{PRIOR} on the evidence one acquires upon opening one's eyes and seeing the minute hand of the clock. P_{PRIOR} satisfies the following (for any i and j):

$$\begin{aligned} P_{\text{PRIOR}}(w_i^i) &= P_{\text{PRIOR}}(w_j^j) \\ P_{\text{PRIOR}}(w_i^{i-1}) &= P_{\text{PRIOR}}(w_i^{i+1}) = P_{\text{PRIOR}}(w_{i+1}^{i+2}) \\ P_{\text{PRIOR}}(w_i^i) &> P_{\text{PRIOR}}(w_i^{i-1}) \text{ (and hence, } P_{\text{PRIOR}}(w_i^i) > P_{\text{PRIOR}}(w_i^{i+1}) \text{)} \end{aligned}$$

The accessibility relation R is such that for any worlds w_i^j and $w_{i^*}^{j^*}$, $w_i^j R w_{i^*}^{j^*}$ just in case $|i - i^*| \leq 1$. Think of the set of worlds accessible from a world w_i^j as the evidence one has in a case in which one is looking at the minute hand of the clock, and it in fact points to i minutes past the hour. The rational credence function at a world w_i^j is gotten by conditionalizing P_{PRIOR} on the set of accessible worlds. For instance,

$$P_{20}(-) = P_{\text{PRIOR}}(-|\{w_{19}^{18}, w_{19}^{19}, w_{19}^{20}, w_{20}^{19}, w_{20}^{20}, w_{20}^{21}, w_{21}^{20}, w_{21}^{21}, w_{21}^{22}\})$$

Figure 1 represents part of the model described:

...	w_{18}^{19}	w_{19}^{20}	w_{20}^{21}	w_{21}^{22}	w_{22}^{23}	...
...	w_{18}^{18}	w_{19}^{19}	w_{20}^{20}	w_{21}^{21}	w_{22}^{22}	...
...	w_{18}^{17}	w_{19}^{18}	w_{20}^{19}	w_{21}^{20}	w_{22}^{21}	...

Figure 1

Since we are assuming that in the actual case the hand points to 20 past the hour, and that the subject's credence function is rational, w_{20}^{20} is the actual case. The 9 worlds that are epistemically accessible from w_{20}^{20} are in bold. The horizontal set of worlds $\{w_{19}^{19}, w_{20}^{20}, w_{21}^{21}\}$ is the proposition that one's credence is rational (whatever the rational credence is). The vertical set of worlds $\{w_{20}^{19}, w_{20}^{20}, w_{20}^{21}\}$ is proposition p_{20} – that the minute hand points to 20 past the hour (and more generally, $p_i = \{w_i^{i-1}, w_i^i, w_i^{i+1}\}$). This is also the proposition that P_{20} is the rational credence function. It is not surprising, then, that these propositions are strongly dependent, since in the model described they are one and the same proposition. The diagonal set of worlds $\{w_{21}^{20}, w_{20}^{20}, w_{19}^{20}\}$ is the proposition that one's credence function is P_{20} (and more generally, $cre=P_i = \{w_{i-1}^i, w_i^i, w_{i+1}^i\}$). In the model described, for any world w_i^i , the credence function that is rational at w_i^i satisfies assumptions 1.-5. above.

Hence, in the model described, evidence about one's credences (that is, evidence stating, of some P_i , that one's credence function is P_i) is neither first- nor higher-order inert. Note also that, as pointed out in the main text, conditionalizing on information about one's credence function does not make it likelier that one's credence function (whatever it is) is rational.

Bibliography

Bronfman, Aaron

2009 *Coping with Imperfection*, Doctoral dissertation.

Christensen, D.

2007 "Epistemic Self-Respect", *Proceedings of the Aristotelian Society* 107: 319-337.

2010 "Rational Reflection", *Philosophical Perspectives* 24: 121-140

2011 "Disagreement, Question-Begging, and Epistemic Self-Criticism", *Philosophers' Imprint* 11:6.

Egan, Andy, and Elga, Adam

2005 "I can't believe I'm stupid", *Philosophical Perspectives*, 19.1: 77-93.

Elga, A.

2007 "Reflection and Disagreement"

2013 "The puzzle of the unmarked clock and the new rational reflection principle", *Philosophical Studies*, 164.1: 127-39.

Evans, G.

1982 *The Varieties of Reference*, John McDowell (ed.), New York: Clarendon Press; Oxford University Press.

Fernandez, J.

2003 "Privileged Access Naturalized", *Philosophical Quarterly* 53:212, pp. 352-372.

Huemer, Michael

2007 "Compassionate Phenomenal Conservatism", *Philosophy and Phenomenological Research* 74: 30-55.

Kelly, T.

2005 "The Epistemic Significance of Disagreement", *Oxford Studies in Epistemology*, John Hawthorne & Tamar Szabó Gendler (eds.), Vol 1, pp. 167-196, Oxford University Press.

Lasonen-Aarnio, M.

2015 "New Rational Reflection and Internalism about Rationality", *Oxford Studies in Epistemology*, John Hawthorne & Tamar Szabó Gendler (eds.), Vol. 5, pp. 145-171, Oxford University Press.

Roush, S.

2009 "Second guessing: A self-help manual", *Episteme* 6.3:251-268

Sorensen, R.

1987 "Anti-expertise, instability, and rational choice", *Australasian Journal of Philosophy* 65.3: 301-315.

Van Fraassen, B.

1984 "Belief and the Will", *Journal of Philosophy* 81: 235-256.

White, R.

2006 "Problems for Dogmatism", *Philosophical Studies* 131.3: 525-557.

2009 "On Treating Oneself and Others as Thermometers", *Episteme* 6.3: 233-250.

Williamson, T.

2014 "Very Improbable Knowing", *Synthese* 79.5: 971-999.