

Predictably Misleading Evidence

Evidence can be misleading: it can rationalize raising one's confidence in false propositions, and lowering one's confidence in the truth. But can evidence be *predictably* misleading? Can a rational agent with some total body of evidence *know* that it points to (a particular) falsehood? It seems not: plausibly, rational agents believe what their evidence supports. Suppose for reductio that a rational agent can see ahead of time that her evidence is likely to point towards a false belief. Since she is rational, if she can anticipate that her evidence is misleading, then it seems she should avoid being misled. But then she won't believe what her evidence supports after all. That is to say, if evidence were *predictably* misleading, it wouldn't be misleading in the first place. So, it seems, evidence cannot be predictably misleading.

The argument sketched above has a lot of intuitive appeal. But it poses a problem for another compelling epistemological view: the view that so-called "higher-order" evidence can require us to revise our beliefs. As I will argue, higher-order evidence is predictably misleading. Insofar as higher-order evidence rationalizes changing one's beliefs, this change tends to result in less accurate beliefs – and we can know this on the basis of an a priori argument. This gives us a new and powerful reason to worry about the significance of higher-order evidence. In this paper I will develop and examine this objection, and explore some possibilities for addressing it.

1. Higher-order evidence

Higher-order evidence, as I'll understand it, is evidence that bears on the functioning of one's rational faculties, or on the significance of other evidence that one has. The following is an example of what I have in mind:

Cilantro. Sam's trustworthy roommate leaves a Tupperware container of chicken curry in the fridge. A sticky note on top reads: "If the following logical proof is valid, the green specks in this curry are cilantro. If not, they are mint." A valid proof follows. Sam (sadly) hates cilantro – it tastes soapy to him – so he works through the proof before risking a bite. He correctly ascertains that the proof is valid and concludes that C: the curry contains cilantro. Then Sam notices a headline in the newspaper: "Gene Causing Soapy Cilantro Taste Linked to Poor Logical Reasoning Abilities". The story goes on to detail the results of a study

showing that people with Sam’s aversion to cilantro perform abnormally poorly on exactly the types of logical reasoning required to assess the proof on the sticky note.

Let’s call Sam’s first-order evidence (the sticky note, etc.) “E”, and his higher-order evidence (the newspaper story) “HOE”. Many people have the following intuitive reaction to cases like **cilantro**:

After examining E, Sam should be highly confident in C. After examining both E + HOE, Sam should reduce confidence in C.

Much recent literature has defended this intuitive reaction, examining epistemic principles that could explain it and emphasizing the odd consequences of rejecting it.¹ But there has also been much debate over how this intuitive reaction could possibly be correct. Some epistemologists suggest that if Sam *does* reduce confidence in C, he is throwing away or ignoring his first-order evidence;² he may be violating plausible epistemic norms like consistency or probabilistic coherence,³ and it is hard to see what kind of “epistemic rule” could be guiding him.⁴ A common thread in the higher-order evidence literature seems to be that there is *something* wrong with Sam if he does not reduce confidence in C, but that it is hard to accommodate this thought in a single, consistent picture of epistemic rationality.

A number of authors – including those who take higher-order evidence to have rational import – have remarked that higher-order evidence seems different from first-order evidence, in important ways. I take the discussion that follows to be in line with these thoughts. The present challenge, however, develops the problem of higher-order evidence in an especially conspicuous and troubling way.

2. The problem for HOE: it is systematically, predictably misleading.

I’ll take as a working assumption that HOE *does* make it rational to reduce confidence in C, as the intuitive reaction suggests. Given that assumption, it’s easy to see how HOE is misleading in our paradigm case, **cilantro**. (What I mean is that it is misleading regarding

¹ See, for example, Christensen [2007], [2010], [2011] and elsewhere; Elga [2007]; and Feldman [2007].

² Kelly [2010], p. 123; Christensen [2009].

³ Christensen [2007] and elsewhere.

⁴ Lasonen-Aarnio [2014].

C, the proposition for which it has distinctly higher-order import.) Sam's first-order evidence indicated that the curry contained cilantro, which, given the trustworthiness of his roommate, is highly likely to be correct. But after reading the newspaper, Sam became much less confident that the curry contained cilantro; his overall belief state moved *farther* from the truth. We can even suppose that Sam's new, lower level of confidence in C rationalized eating some of the curry; in that case it looks like respecting HOE was a real tragedy for Sam!

Sam's situation isn't just an unlucky feature of this particular case, **cilantro**. Here is a more general argument:

Higher-order evidence is misleading

- | | |
|------------|--|
| Premise 1 | First-order evidence tends to point towards the truth. |
| Premise 2 | Higher-order evidence tends to point away from what the first-order evidence supports. |
| Conclusion | Higher-order evidence tends to point away from the truth. |

As I will argue here, the conclusion of this argument is a true generalization about higher-order evidence. Higher-order evidence tends to be misleading. And since the argument above is made on a priori grounds – P1 and P2 simply follow from reflections on the nature of evidence and how it works – higher-order evidence is *predictably* misleading. Indeed, Sam himself could go through this argument and figure out that *his own* higher-order evidence is likely misleading.⁵

Before defending the premises, I would like to pause to clarify a few things: two about the argument's terminology, and another about the form of the argument itself. First, the phrase "tends to" in this argument is deliberately loose and informal. There are a number of ways in which one could make it more precise. (For example, for the first premise, one might say: "when the first-order evidence supports P to degree n, P is true in n% of nearby possible worlds". Or one might say: "The evidence supports P if it is a reliable indicator of P." Or one might say: "when the first-order evidence supports P

⁵ Note that I'm not making this stronger claim: that Sam can know that his higher-order evidence is *definitely* misleading. What he knows is that it is *probably* or *likely to be* misleading, just as in ordinary cases we know that our evidence is *probably not* misleading.

highly, the expected accuracy of P is high.”⁶ And so on.) For present purposes, I won’t endorse any particular one of these.⁷

A second terminological note concerns what it means to say that higher-order evidence is misleading. As I understand it here, a piece of evidence is not just misleading, full stop, but misleading *with respect to a proposition, in an evidential context*. As I mentioned above, I am interested in whether higher-order evidence is misleading with respect to those propositions for which it has distinctly higher-order import. In **Cilantro**, it is natural to talk about *the newspaper story*, or some proposition concerning the newspaper story, as a piece of evidence. That piece of evidence rationally affects Sam’s attitudes about a number of different topics, in a number of different ways. For example, it might rationally raise his confidence *that some scientists are studying the cilantro-soap-taste gene, or that the soap-taste gene is connected to logical ability*. A piece of evidence might be misleading regarding some of these propositions and not others.⁸ Here, I am most interested in the way higher-order evidence affects the proposition(s) that it targets *qua higher-order evidence*. The second qualifier, “in an evidential context”, is necessary for the familiar reason that the import of a piece of evidence can change based on what other evidence one has. A piece of evidence can be misleading against the backdrop of one body of evidence, but not another. Here, I am interested in a particular backdrop as well: the accompanying first-order evidence. So when I say that higher-order evidence is misleading, what I mean is that *against the backdrop of its accompanying first-order evidence*, higher-order evidence tends to make one’s rational opinion *about the proposition(s) that it targets qua higher-order evidence* less accurate. (Again: all of this is just to articulate the claim I will defend. I haven’t yet defended it!)

Finally, note that this argument is not deductively valid. (I formalize it here not to show a logical implication, but to isolate the assumptions that support the claim I am

⁶ See Schoenfield [2015 a] for a similar discussion, put in terms of expected accuracy. There, Schoenfield argues that respecting higher-order evidence does not maximize expected accuracy.

⁷ In presenting this argument I have sometimes gotten the objection that the argument is nonsensical without a more precise understanding of “tends to”. However, I think the objection gets things backwards: however we eventually want to understand the relationship between evidence and truth, we should be constrained by something like P1. If P1 turns out to be false, our theory is probably wrong.

⁸ Because evidence has many effects all at once, it might be more accurate to talk about higher-order rational effects, rather than higher-order evidence. To stay in line with the literature, I will discuss “higher-order evidence” as if it is a separate kind of evidence; however, readers can keep in mind that I am only focusing on the distinctively higher-order impact of this evidence.

interested in.) For it could be that we *only*, or *almost always*, receive higher-order evidence in those rare situations where the first-order evidence is misleading. This is the case with undercutting or rebutting defeaters: we tend to get evidence that the lighting is tricky when it *is* tricky, for instance, and the jellybean isn't red after all. In fact, one might think that ordinary defeat works precisely by alerting us to the fact that we are in an odd situation where our first-order evidence is misleading.⁹ But this is not how things are for higher-order evidence. There is no reason to think that when we encounter higher-order evidence should be at all correlated, positively or negatively, with whether our first-order evidence is misleading. Misleading evidence is hard to spot and comes along more or less randomly (unless someone is trying to trick us – but that is not correlated with receiving higher-order evidence). Therefore, since higher-order evidence tends to point away from what the first-order evidence supports, we should expect that higher-order evidence tends to point away from the truth.¹⁰

This looks bad. Plausibly, epistemic rationality involves believing what one takes to be true, from one's own perspective. Epistemic rationality also, plausibly, involves believing what one's evidence supports. Cases like **cilantro**, which appear to be cases of predictably misleading evidence, bring out a tension between these plausible thoughts.

9 Allan Coates [2012] makes a similar point regarding this difference between higher-order defeat and ordinary defeat: ordinary defeat indicates that one's earlier belief (that the jelly bean is red, for example) *was* justified, but that one would no longer be justified in holding it (now that one knows about the tricky light). Higher-order defeat, on the other hand, suggests that one's earlier belief was not justified to begin with.

¹⁰ This point highlights an important difference between higher-order defeat and other kinds of defeat. As I mentioned, we tend to encounter undermining defeat when our first-order evidence is misleading – it is what tips us off to this fact about our evidence. But higher-order evidence is different. Like undercutting or rebutting defeat, it counters the effect of other evidence that we have; unlike these other kinds of defeat, however, it does *not* indicate that the evidence it targets is misleading.

Noticing this difference also helps preempt an objection I have encountered from several people. The objection says: couldn't we make a parallel argument about any old *new* piece of evidence? After all, new evidence often makes it rational to change our beliefs, so by this argument, isn't "the second piece of evidence" predictably misleading? This argument sounds fallacious, for much the same reason that Kripke's dogmatism argument sounds fallacious.

However, I think my argument here is different from this fallacious argument. The "second piece of evidence" can have two kinds of impact: it can go against what the first-order evidence supports, or it can agree with the first-order evidence and point in the same direction. We have already discussed why a parallel argument won't work in the first case, regarding rebutting and undermining defeaters. Although the analogue of P2 is true, we also have reason to think that we will generally encounter rebutting or undermining defeat in cases where the first-order evidence is misleading. And in the second case, the analogue of P2 is not true. New, non-undermining evidence does not tend to point away from what the rest of one's evidence supported. So it seems to me that the argument here is unique to higher-order evidence.

Thanks to [OMITTED].

How could it be rational to believe what our evidence supports, if we know that doing so is likely to lead us *away* from the truth?

Before jumping into that challenge, though, let me say more to motivate it. I will defend P1 and P2 in the following two subsections.

2.1 Defending P1

First, here is P1 again:

P1 First-order evidence tends to support the truth.

I do not have much to say in defense of P1. Some skeptics might deny it. But anyone who thinks that evidence is relevant to what we should believe, and often rationalizes changing our opinions about the world – anyone who herself strives to believe what her evidence supports, and revises her beliefs on its basis – should accept P1 on pain of incoherence. After all, what we believe is what we take to be true. It is incoherent to set out trying to believe something we take to be false. Since we expect evidence in general to support the truth, we should also expect first-order evidence to support the truth.¹¹

2.2 Defending P2

I take P2 to be the more controversial premise in the argument above. Here it is again:

P2 Higher-order evidence tends to point away from what the first-order evidence supports.

I will explain what P2 means first, and then why we should hold it. To rephrase P2: if one's first-order evidence rationalizes a certain change in one's beliefs, then higher-order evidence will tend to rationalize making a change in the opposite direction. For example, if one's total first-order evidence makes it rational to increase confidence in P (relative to one's prior confidence in P, before receiving that evidence), then higher-order evidence bearing on that first-order evidence will tend to make it rational to decrease confidence in P (relative to one's prior confidence in P given the first-order evidence alone).

¹¹ Should we be suspicious about this motivation, given that *first-order* evidence is only a subset of our total evidence? I don't think so – not unless we have some reason to think that the first-order evidence is a particularly unrepresentative or misleading sample of our total evidence. I take it we also have (prima facie) reason to accept that evidence collected on Tuesdays tends to support the truth, evidence about penguins tends to support the truth, etc.

Again, it is important to note that in P2, I am focusing on the propositions that higher-order evidence targets *qua higher-order evidence*. In this case, that means Sam's belief about C: whether the curry contains cilantro. My claim in P2 is that, insofar as higher-order evidence has a distinctive sort of impact, it works by defeating, wiping out, neutralizing, or weakening the effect of the first-order evidence that it targets.

In **cilantro**, Sam's first-order and higher-order evidence point in opposite directions. Sam's first-order reasoning supports P, and after going through the proof it is rational for Sam to become highly confident in P. But after going through the proof *and reading the newspaper*, it is rational for Sam to become less confident in P. So Sam's higher-order evidence counteracts the effect of his first-order evidence, concerning C.

To argue that this phenomenon is more general – it is not just a special feature of **cilantro** – I will look at two objections to P2. Both of these objections say, in different ways, that the phenomenon *is* a special feature of **cilantro**. I will argue that these objections fail.

(A) Objection 1: higher-order confirmation?

P2 says that higher-order evidence tends to undo or counteract the effect of first-order evidence. Let's call this phenomenon "higher-order defeat." (**Cilantro** is a case of higher-order defeat, as are most of the cases of higher-order evidence discussed in the literature.) But in some cases, higher-order evidence instead *confirms* what the first-order evidence supports. Call that phenomenon "higher-order confirmation". So one might object as follows: *There's no reason to think that higher-order defeat is more common is than higher-order confirmation. Therefore, we have no reason to believe P2. In fact, we can't say anything general about what rational effect higher-order evidence "tends" to have, relative to what the first-order evidence supports. Sometimes it defeats, and sometimes it confirms.*

I agree that higher-order confirmation is possible. Here is an example:

Peer agreement: We all take a test, and I'm not sure that I got the right answer to question 5. In fact, I did get it right, and I am rationally pretty confident in the right answer – though I do have some doubt, so I'm not completely confident. Then I talk to the rest of the class, afterwards, and find out that everyone else got the same answer that I did.

My classmates' agreement is higher-order evidence: it bears on the significance of my first-order evidence, and suggests that I did in fact accommodate it rationally. And intuitively, it should make me *more* confident in the right answer than I was before. It therefore points in the *same* direction as the first-order evidence.¹²

However, I don't think that the possibility of higher-order confirmation removes, or even mitigates, the central challenge. First, even if we agree with the objector that higher-order confirmation is just as common as higher-order defeat, we could reformulate the challenge by restricting our attention to cases of higher-order defeat. Higher-order *defeating* evidence still tends to be misleading (because it defeats first-order evidence, which tends to point to the truth). And higher-order defeating evidence is still *predictably* misleading. After all, Sam can tell that his higher-order evidence is defeating, not confirming: so, he can predict that it is misleading in his case. (Sam's higher-order evidence suggests that he has accommodated his first-order evidence *badly*, not well, and he knows that.) So higher-order evidence is still predictably misleading in paradigm cases like **cilantro**, and the problem remains for agents like Sam.

I think that there is a stronger response available, as well. That is: even though higher-order evidence can provide confirmation, it is still true that higher-order evidence is predictably misleading in a very important sense. That is: insofar as one's total higher-order evidence rationalizes *changing* one's opinion – that is, insofar as it rationalizes holding a *different* opinion than what would be rational *without* it – it tends to be misleading. Insofar as higher-order evidence makes a *net difference* to what's rational for you to believe, it's for the worse.

To illustrate the point, let's look more carefully at cases like **peer agreement**. In that case, I started off with some doubt about my rational abilities, which is why I was not entirely confident in my answer. The higher-order evidence I received from my classmates *counteracted* the higher-order doubt that I had previously. So there is one piece of higher-order evidence that points to the truth in this case – but it is only able to do so against the backdrop of my prior higher-order doubt.¹³ In **peer agreement**, my

¹² My case is based on one presented in Christensen [2011].

¹³ If you're not convinced, compare this case to another one, in which I go into the math test with *no* reason for higher-order doubt – I am simply not aware that people tend to make mathematical mistakes, and have no knowledge of my own or my peers' track record. Call this case "**peer agreement with no prior HOE**".

belief about the answer to question 5 becomes *more accurate* only because the net effect of my higher-order reasoning gets *smaller*. This is very strange: with first-order evidence, we tend to become more accurate as the net effect of that evidence gets *larger*.¹⁴

(B) Objection 2: higher-order evidence for irrational agents?

Another way to deny P2 focuses not on different kinds of *cases*, like peer agreement, but on different kinds of *people*. This objection says: *Higher-order evidence is misleading for people like Sam, in cases like cilantro. However, there are other people for whom the very same higher-order evidence would not be misleading in cases like cilantro. Higher-order evidence therefore sometimes points in the same direction as first-order evidence, and sometimes points in the opposite direction. This is true even given the very same body of first-order evidence and higher-order evidence.*

To see what kind of person this objector has in mind, suppose Sam has another roommate, Sally. Sally is also sadly afflicted by a dislike of cilantro. Suppose Sally is now in just the same circumstances as Sam is in: she sees the curry in the fridge, reads the note, and completes the proof. But where Sam correctly concluded that C (the curry contains cilantro), Sally makes a logical mistake, judges the proof to be invalid, and concludes $\sim C$. (Perhaps this is because Sally, unlike Sam, *is* compromised in her logical

Unrealistic as it is, suppose that this could happen. If I had no higher-order doubt at all, what should I believe about the answer to question 5? A natural answer is that I should believe exactly what my evidence supports. So I should be highly confident of the answer. Now suppose that, just as in the original **peer agreement** story, I talk to everyone else after class and find that they put down the same answer. Should I become more confident in the answer I gave? It is hard to see why my credence should change at all in this case. This suggests that higher-order confirmation (in the sense of raising one's confidence) in general can only happen against a backdrop of higher-order defeat.

¹⁴ An interesting consequence of this discussion is that, plausibly, as a *rational* agent acquires more and more higher-order evidence, the net effect of her higher-order evidence *will* get smaller and smaller. (Facts about our own rationality are first-order facts about the world, and we should expect that as we learn more about the world, a rational agent's views about her own rationality will become more and more accurate. And since she *is* rational, as she gains more higher-order evidence, she will come to have less and less reason to *doubt* that she is rational.) This observation might initially seem to alleviate our central worry. But on the contrary, I think it only makes the worry more pressing. In general, if we know that gaining more and more evidence will eventually have some particular epistemic effect, we don't need to wait for the evidence to come in; we can just believe what it will support. (This sort of thought is what's behind van Fraassen's Reflection principle, Feldman's "evidence of evidence is evidence", and other related principles.) In this case, we know that more and more higher-order evidence will have less and less of an effect on a rational agent's first-order beliefs. So why should higher-order evidence have any effect at all? Why not believe what we know a *larger* body of higher-order evidence would support – that is, why not believe what we should believe if higher-order evidence had *no effect at all*? Why not just believe exactly what our first-order evidence supports? Thanks to [OMITTED] for raising this point.

reasoning abilities due to the cilantro-soapy-taste gene.) Then Sally reads the story in the newspaper, and becomes worried about the reasoning she has just completed. For her, the effect of the newspaper story is to reduce her confidence in $\sim C$, thereby increasing confidence in C . Sally's situation is therefore one in which *higher-order defeat* leads to a *more accurate* first-order belief, and a first-order belief that is more in line with what the first-order evidence supports. So, according to this second objector: in Sally's case, the higher-order evidence *points to C – that is, it points in the same direction as the first-order evidence*. Whether higher-order evidence agrees or disagrees with first-order evidence depends on whether you are like Sam, or like Sally. Therefore, P2 is false.¹⁵

There is a lot to say about Sally's case, and how it should be explained within our larger theory of epistemic rationality.¹⁶ Here, I will just point out two related reasons to resist the present objection. As before, I will assume that Sam's higher-order evidence really does *support* his lowering confidence in C , and hence his total evidence at the end of the story supports low confidence in C . The issue here will be whether we can say the same thing about Sally.

First: as several people have noted, there seems to be an asymmetry between agents like Sally and agents like Sam. At the end of the story, Sam seems *more rational* than Sally, even though they both did something procedurally sensible in response to their higher-order evidence.¹⁷ The present objection, which treats the two cases as equals, will have a hard time explaining the sense in which Sam's final belief state is more rational.

Second, and more importantly, the objection relies on a certain interpretation of “where the evidence points”, which I will argue is mistaken. This objector seems to want to say that the evidence points one way for Sally and another way for Sam, despite Sally and Sam *sharing* their evidence. So in this case, the very same evidence points to both truth *and* falsity. On the interpretation that I favor, Sam and Sally's total evidence

¹⁵ Furthermore, this argument seems to have the resources to say that, not only is it false that higher-order evidence tends to be misleading, but in fact it is true that higher-order evidence tends to point to the truth. Actual, fallible agents who receive HOE are in Sally's position more often than we're in Sam's. That means that for us, higher-order evidence will tend to point towards the truth. Thanks to [OMITTED] for very helpful discussion here.

¹⁶ For further discussion, see Horowitz & Sliwa [2015] and Schoenfield [2015 b] and Christensen [2016 a]. Also see Kelly [2010] and Christensen [2011] for an exchange on this point re: peer disagreement.

¹⁷ See Christensen [2011], Kelly [2010], Horowitz & Sliwa [2015].

supports confidence in C, despite the fact that Sally ends up with high confidence in C after responding to that evidence. If my interpretation is right, it explains the asymmetry between Sam and Sally (Sam is more rational because his beliefs align with where the evidence points, or what it supports, while Sally's beliefs do not) and also explains why the present objection is on the wrong track. Let me explain what I mean.

In order to know where a certain sign or signal points *to*, we often need to know where it is pointing *from*. Different kinds of indications or signs, occurring in different contexts, lend themselves to different interpretations. One example: on a hike, you encounter a blue arrow painted on a tree, pointing up the mountain. This arrow is pointing *from* your present location, exactly where you are standing when you can see the arrow. The arrow's meaning is something like, "from *here*, go up the mountain to follow the blue trail." A different kind of example: you are an explorer on Treasure Island, following directions to a pirate's hidden chest of jewels. Step 5 of the pirate's instructions says, "Turn left and walk 30 paces. The treasure is buried here." Obviously, this step of the direction is not pointing *from* one's present location unless one has also correctly followed Steps 1-4. Instead, Step 5 points to some location *relative to the preceding steps*.

An important difference between these two examples comes out in cases where we fail to reach the destination. In the first example, suppose you followed the blue marker up the mountain and found yourself on the green trail rather than the blue trail. This is the trail marker's fault; it pointed in the wrong direction. Now take the second example: suppose you turn left, walk thirty paces, and start digging. You find no treasure. Does this mean that the pirate's instructions were wrong? Well, that depends: did you follow Steps 1-4? If you did, then we can blame the pirate's instructions. But if instead, you wandered randomly around the island, checked the directions, and jumped straight to Step 5, then it's your own fault that you did not find the treasure.

Returning to epistemology, we might understand *evidential* pointing in either of these two ways. If a piece of evidence is like a trail marker, "pointing" an agent from wherever she happens to be, it would be apt to describe Sally's evidence as "pointing to the truth". Sally responded to her higher-order evidence given the (not fully rational) beliefs she had at the time, and ended up closer to the truth about C. However, if a piece

of evidence is more like a step in the pirate's directions, this description would not be apt. On this second understanding, a piece of evidence points to one belief or another relative to the rest of one's total evidence – not relative to one's current "epistemic location". While Sally may have done something right in response to her higher-order evidence, it is not accurate to say that higher-order evidence pointed to the truth in her case.

Merely drawing this distinction, I think, puts pressure on the present objection: it is not obvious that Sally's evidence is non-misleading, simply because it's not obvious that we should interpret "evidential pointing" in the first way rather than the second. Moreover, I am inclined to think that the second interpretation of "evidential pointing" is the correct one: we should see it as more like the pirate's directions than like the trail marker. When we talk about whether a body of evidence is misleading, we care about which of our epistemic failures or successes can be credited to *the evidence itself*, rather than to the person assessing that evidence. So want to know whether a bit of evidence points to the truth *relative to the rest of a total body of evidence*.¹⁸

2.3 Summing up

I have claimed that we should accept both P1 and P2. First-order evidence tends to point to the truth, and higher-order evidence tends to point in the opposite direction from the first-order evidence that it targets. We should also expect higher-order evidence to crop up at any time – not only when the first-order evidence is misleading. So it seems to me that we should accept the claim that higher-order evidence tends to be misleading. But if higher-order evidence predictably points away from the truth, why should we care about believing what it supports? And how can higher-order evidence be evidence at all?

¹⁸ One might get the impression that, in arguing against this objection, I am begging the question against permissivism. (That's because the objection claims that the evidence can point in two directions at once, and my reply denies this.) However, I hope that the map/trail marker analogy makes it clear that I am not. Most permissive views of rationality hold that different agents have different treasure maps – whereas the objector I am responding to here thinks that evidence is not like a treasure map at all, but like a trail marker.

Notice that even if we deny that Sally's change of belief reflects "where her evidence points", we can still make use of other forms of positive epistemic appraisal to describe Sally's reasoning in response to the newspaper story. For instance, James Pryor's notion of "rational commitment" (Pryor [2004], pp. 363-365) might be relevant. Rational commitment, unlike evidential support, is relative to one's prior belief states. Pryor writes that someone who believes "Johnny can fly" is rationally committed to believing "someone can fly", even if the belief that Johnny can fly is completely unsupported by this person's evidence. Believing what one is rationally committed to might be praiseworthy in some sense, even though rational commitment can come apart from evidential support – and can come apart from what it is all things considered rational for one to believe.

This is the challenge for views according to which higher-order evidence has a rational effect. In the remainder of the paper I will look at some particular views about how higher-order evidence works, asking whether each view has the resources to address this challenge. I'll argue that the first two do not, but the third shows some promise.

3. Two-norm solutions

In the first half of the paper I posed the challenge for higher-order evidence: if it has a rational effect, it is predictably misleading. But evidence should be predictably truth-guiding, not predictably misleading. So the intuitive thought that higher-order evidence *does* have an effect is inconsistent with the intuitive thought that evidence should guide us to the truth.

The first sort of view I'll discuss embraces both of these contradictory claims. This sort of view says that higher-order evidence *is* rationally significant, *it is* predictably misleading, and yet evidence *does* lead to the truth. We can hold these claims all at once because, on this view, there are two different senses of “rationality” or “what we should believe” at work in cases of higher order evidence. In one sense of “rational” – the one that corresponds to evidential support – it is rational to *ignore* higher-order evidence. It is this first sense of “rational” that makes the following sentence true: “Rationality is a guide to the truth.” But in another, derivative sense of “rational”, it is rational to respect higher-order evidence. I'll call views of this style “two-norm views”.

The idea that our apparently inconsistent normative judgments track two (or more) norms is familiar from ethics, where philosophers often acknowledge a difference between “subjective” and “objective” moral norms or “oughts”. In the current literature on higher-order evidence, epistemologists have drawn two-norm distinctions in a number of ways: (a) objective vs. subjective; (b) reasons vs. rationality; (c) an evidential norm vs. the dispositions that generally lead one to follow it; (d) the norm one should follow vs. the norm that one should *try* to follow; (e) the “best method” vs. the best method to adopt; (f) the plan you should conform to vs. the plan you should *make*.¹⁹ In many of

¹⁹ See Sepielli [ms] for (a); Worsnip [2018] for (b); Lasonen-Aarnio [2010] for something close to (c) and for (e); Williamson [2000] for something close to (d); and Schoenfield [2015 a] for (f). Lasonen-Aarnio [ms] speaks more generally of the evidential norm vs. a “derivative norm”, with the details of how derivative norms are derived left for later.

these cases, one norm is understood as “ideal” and the other as “non-ideal”. The upshot of all of these proposals, of course, is that the *first* norm in the pair tells Sam to maintain high confidence in C. The *second* norm tells him to reduce confidence in C in response to his higher-order evidence.

3.1 A challenge for two-normers

Two-norm solutions are somewhat unsatisfying for one obvious reason: they leave open the question of *what we should believe*. There is therefore a simple argument against these views: since epistemology’s primary aim is to answer this question, two-norm views have failed.²⁰ I am sympathetic to this simple argument, but it is a bit too quick; after all, perhaps these views are right to say that we have been up to now confused about the aim of epistemology, and that the question of what we should believe is ambiguous or equivocal. Indeed, the fact that we seem to have run into a contradiction in answering the question is good evidence that it *is* equivocal. So I would like to focus instead on a related challenge. What is it about these two norms that makes them both *distinctly epistemic*, and both *worth caring about*? To answer this challenge, I take it that defenders of two-norm views need to explain how each norm is related to the pursuit of truth. If some mode of rational evaluation is *not* connected to truth, it is either not distinctively epistemic or else it is not worth caring about.

Some two-normers might readily admit that this challenge can’t be met. For such people, perhaps only one of the norms is meant to be interesting or worth caring about in the first place. One gets this sense from, for example, Timothy Williamson, who often writes about various epistemic rules being the right ones to “try to follow” (as opposed to being the right rules to follow), but does not make any attempt to unify this more subjective category forms anything like a theoretically useful or coherent whole.²¹ Or perhaps one of the norms is epistemic, and the other is practical.²² However, I am currently interested in the prospects for *solving* (rather than giving up on) the problem at hand. So I will focus here on the possibility that there are two incompatible, important, distinctively epistemic norms.

²⁰ Kvanvig [2014], ch. 3, develops this view in detail.

²¹ Williamson [2000]. See, for example, See also Weatherson [2014] for a similar view about morality.

²² Smithies [ms] develops this view.

It seems to me that the most promising two-norm approach should say that both norms are connected to the truth, but in different ways. This is how I interpret some of the suggestions above, particularly those that draw the distinction along ideal/non-ideal lines. What is good advice for ideal agents might be terrible advice for non-ideal agents, and vice versa. So, if we think of epistemic norms as giving guidance or advice, it is plausible that we would end up with different norms for different kinds of agents. Schoenfield [forthcoming] develops this approach ((f), in the list above), specifically addressing the connection between higher-order evidence and truth or accuracy. I will focus on her view here.

Schoenfield argues that when we are making epistemic plans, or asking for epistemic guidance, we should consider the effect of making one plan or another. We should consider the *effect* because, in Schoenfield's view, epistemic rationality aims to achieve certain goals (namely, accuracy). So what's relevant to assessing a plan is how well it achieves those goals. We should consider the effect of *making* a plan because we are fallible, and sometimes don't follow through.

To illustrate the distinction between the best plan to conform to and the best plan to make, Schoenfield imagines a chronically late person who plans to leave the house at 10:00 so that she will *really* leave at 10:30.²³ Leaving at 10:30 may be the best plan to conform to; but if you are chronically late, leaving the house at 10:00 is the best plan to *make*. Schoenfield argues that fallible believers such as ourselves (or more carefully, believers who have reason to think they are fallible) are in a similar position with respect to our "epistemic plans". We should make plans – that is, hold ourselves to epistemic norms – that will yield the best results for fallible agents like us, who can at best hope to follow these plans imperfectly.

Schoenfield argues that this distinction can help us make sense of cases involving higher-order evidence. In cases like **cilantro**, Schoenfield's view is that the best plan to *conform to* is to believe what your first-order evidence supports, ignoring higher-order evidence. But since we can't expect to follow that plan perfectly – after all, we know that

²³ Schoenfield [2015 a], p. 651. See her sections 4 and 5 for discussion of how this distinction bears on higher-order evidence in general.

our reasoning might be impaired in various ways – we should *make* the plan to respect higher-order evidence as well.

Why should we agree that the best plan to conform to has us ignore higher-order evidence? Schoenfield’s argument is in line with the argument I have given above: higher-order evidence is predictably misleading. Schoenfield imagines a “perfectly rational ignorant agent” who is programming an infallible robot to form opinions in response to evidence. If this agent only cares about her robot’s accuracy, she will program the robot to ignore higher-order evidence. That’s because, in cases like **cilantro**, respecting higher-order evidence would result in the robot having *less* accurate beliefs.²⁴ I agree with Schoenfield’s assessment here.

What about the second part of Schoenfield’s view: the idea that the best plan to *make* involves revising our beliefs in light of higher-order evidence? This is where problems arise.

First of all, it is not obvious that respecting higher-order evidence *is*, in fact, the plan that is best-to-make in Schoenfield’s sense. Recall that a plan is best-to-make by virtue of the consequences of making it, for real agents in the real world. So the question of which plan we should make is an empirical question, which depends on the particulars of our own psychology and the ways in which we are likely to live up to, and fall short of, our plans. These factors are complex and hard to predict; moreover, they vary from agent to agent. A plan to respect higher-order evidence might have the desired effects for some level-headed agents, for example, but might send insecure agents into a tailspin of self-doubt, and might leave some arrogant agents’ irrational beliefs intact. It would be remarkable if there was any one plan that was best-to-make for more than one person, *and* that plan ended up being anything like the general rules for higher-order evidence that have been defended in the literature. Given these worries, it is questionable to what

²⁴ The robot analogy suggests a compelling view of the nature of evidential support: on this view, evidential support is something like “the a priori best program for an infallible belief-forming robot”. This view seems to be fairly common among epistemologists, dating back at least to Carnap; see, for example, his [1966]. Carnap writes that in constructing an inductive logic, we might usefully consider “a robot with organs of perception, data processing, decision making, and acting.” He continues: “Thinking about the design of this robot will help us in finding rules of rationality. Once found, these rules can be applied not only in the construction of a robot but also in advising human beings in their effort to make their decisions as rational as their limited abilities permit.” (p. 265)

extent Schoenfield's suggestion can really vindicate our intuitions in cases like **cilantro**.²⁵

However, even setting this issue aside, Schoenfield's view faces other challenges due to its general structure. In particular, the view's focus on *consequences* leads it to overgeneralize in intuitively problematic ways.

First: although this type of two-norm view focuses on epistemic *consequences* – in this case, consequences for true belief or accurate credences – it is not clear why the *norms* that it supports need to be purely epistemic. For example, I tend to reason badly when I'm hungry. So maybe I should make this plan:

Sandwich Plan: Have a sandwich before engaging in difficult reasoning.

My Sandwich Plan is a good one: I know that I reason better when I'm not hungry, and end up with more accurate beliefs if I have a sandwich first. I'm also quite good at following through on my plans to eat, so the Sandwich Plan is not only a good plan to *conform to*, but also a good plan to *make*.

With due respect to its obvious merits, however, the Sandwich Plan is not the sort of plan we are interested in in the present context. Epistemology should not tell me to have a sandwich. If we really are focused on the *consequences* for true belief, several kinds of inquiry would serve us better: psychology, optometry, etc. More generally, then, it is hard to see why this rationale for a two-norm view yields two distinctly *epistemic* norms, rather than practical ones.²⁶

Schoenfield could rule out the Sandwich Plan by simply stipulating that her view only applies to purely epistemic plans, whose inputs and outputs are purely epistemic (for instance: evidence, beliefs and credences, and inductive standards). A paradigmatic example of a purely epistemic plan might just be a function from total evidence to

²⁵ Schoenfield acknowledges this general type of worry ([2015 a], p. 654-655), and points out that even though we cannot be sure of the *actual* consequences of making any given plan, we can always assess a plan's *expected* consequences using whatever evidence we have. Her view is that the best-to-make plan is the one with the best expected, not the best actual success. This seems right to me, but it does not address my central worry. While it may be true that there is *some* best-to-make plan (for any agent, given her evidence), it is not clear that this plan will include responding to higher-order evidence in any systematic way, or that there will be any epistemic plans that are best-to-make for more than one agent.

²⁶ See Lasonen-Aarnio [forthcoming] for a similar suggestion.

doxastic states. This move would eliminate the Sandwich problem, since the Sandwich Plan has some non-epistemic outputs (“have a sandwich”).

But even if we make this move, other counterintuitive consequences remain. For instance, consider Ivan, who has severe arachnophobia and reasons badly whenever he believes that a scary spider is nearby. Given his particular shortcomings, perhaps this is the best plan for him to make:

Spider Plan: Never, ever believe that there is a spider nearby.

Suppose Ivan could get himself to conform to this plan, training himself to ignore all evidence of nearby spiders. (This might be easier than training himself to get over his arachnophobia!) He would thereby end up with many more true beliefs than he would have, had if he took the spider-evidence into account.

Ivan’s plan looks rational by Schoenfield’s lights, even with a stipulated restriction to purely epistemic plans. (Neither the inputs nor the outputs of the Spider Plan include anything non-epistemic.) While the Spider Plan is a bad plan for infallible rational agents, who wouldn’t be affected by arachnophobia in this way, it’s a good plan for Ivan. It helps him compensate for his own fallibility, and leads to more accurate beliefs in the long run. But this is the wrong result. It is epistemically irrational for Ivan never to believe that there are spiders nearby, even when he can clearly see a spider right in front of him, receives reliable testimony that there is a spider in the vicinity, or hears the distinctive pitter-patter of hairy little legs.²⁷

²⁷ The irrationality of Ivan’s Spider Plan raises a broader potential worry: it suggests that there is a problem with consequentialist views in epistemology more generally. This issue has been an important topic in recent literature, owing to the fact that focus on accuracy in epistemology is often put in consequentialist terms. If accuracy-based views in epistemology must be consequentialist, then Spider Plan problems will crop up all the time. Indeed, one might think that similar worries will even crop up for Schoenfield’s infallible robot. If your only goal in programming the robot is for it to end up with accurate beliefs, shouldn’t you also program it to accept irrational-looking tradeoffs – for instance, if someone offers the robot three true beliefs in exchange for one false one, shouldn’t the robot be programmed to accept? I think that this worry can be addressed, as long as it is possible to think about the importance of accuracy in a non-consequentialist way. (For further discussion of how this might be accomplished, see Carr [2017] and Konek and Levinstein [forthcoming].) But I will leave this line of thought to one side here. In the meantime: while all accuracy-based accounts *might* have a spider problem, explicitly consequentialist accounts like Schoenfield’s *definitely* do.

The upshot of all of this is that two-norm views face a challenge defining the second norm: making it distinctly epistemic, clearly tied to the truth, and plausible.

3.2 Another “two-norm” view: epistemic dilemmas?

Similar problems arise for a different sort of view, which says that in cases like **cilantro**, agents face an “epistemic dilemma”. According to this view, which has been defended by David Christensen, there are two rational requirements that can’t be satisfied at the same time in cases like **cilantro**.²⁸ So no matter what Sam believes, he is guaranteed to violate at least one rational obligation.

The dilemma view is not exactly a two-norm view. It does not say that there are two modes of rational evaluation, and it does not attribute puzzles of higher-order evidence to any kind of equivocation. But it has similar trouble explaining why our epistemic obligations in these cases are genuine obligations, and why we should care about meeting them.

Christensen writes that in cases like **cilantro**, although we can’t perfectly satisfy all of our epistemic obligations at once, there is nevertheless a *maximally rational* response.²⁹ Christensen thinks that the maximally rational response here is to decrease confidence in C. But why is it maximally rational to decrease confidence in C? And if it is, why should we care about being maximally rational? As I’ve argued, if it is rational to revise confidence in response to higher-order evidence, rationality is predictably and systematically misleading. On Christensen’s view, *maximal* rationality is predictably and systematically misleading.

The suggestion that Sam faces a rational dilemma does justice to the thought that, if he responds to higher-order evidence, Sam is doing something *wrong*. But the dilemma view still can’t explain why Sam is also doing something *right* in decreasing confidence in C – and it certainly can’t explain why Sam’s response is *maximally rational*. The original puzzle remains.

4. Calibration views

²⁸ For example, see Christensen [2007], [2013], and [2016 b].

²⁹ If there were no maximally rational response, the dilemma view would be *very* similar to the two-norm views.

So far I have argued that higher-order evidence is misleading, and predictably so. This is a problem: why should we believe what our evidence supports, if we know that it points away from the truth? I looked at one type of radical response we might have to this problem and argued that it was unsatisfactory. So what should we do?

In this final section I will explore the possibility that existing, mainstream accounts of higher-order evidence have the resources to address this challenge. To do so, these accounts will need to draw a distinction between a rational agent's perspective and her body of total evidence. They will also need to deny that it is always rational to believe what our evidence supports.

It might help to make things more explicit. Throughout the previous three sections, I have been working with the following assumptions:

- (1) Evidence tends to point to the truth.
- (2) Rationality requires believing what's likely from one's own perspective.
- (3) Rationality requires believing what's likely given one's evidence.
- (4) Rationality requires revising one's beliefs in light of higher-order evidence.

I also argued for the following:

- (5) Higher-order evidence is misleading: it rationalizes changing one's beliefs in a way that tends to make those beliefs less accurate.

If we accept (5), this seems to create a problem – most likely, it might appear, a problem for claim (4). In sections 1 and 2, I focused on developing this problem. In section 3 I looked at a way we might resolve the problem, by positing an equivocation in what (3) and (4) mean by “rationality requires”.

This section will be focused on a different response, aimed at claims (2) and (3). So far I have not separated these claims. (2) is almost tautologous: to believe something that one takes to be *unlikely* (i.e., to believe something that is unlikely from one's own perspective) seems paradigmatically irrational. And (3) looks, at first glance, like just one way to spell out (2) in more detail. If one's “perspective” just is one's total evidence, then rationality requires believing what is likely given one's total evidence.

My suggestion here will be that (3) is *not* just a more precise statement of (2); and that moreover, we can reject (3). In so doing, we can keep (1), (2), (4), and (5). The resulting view says that higher-order evidence *does* change what's likely from your

perspective, and hence (because we are keeping (2)) what you should believe. But, on this view, higher-order evidence *does not* change what's likely on your total evidence. So higher-order evidence should not be thought of as providing ordinary evidential support.³⁰

4.1 Total evidence and rational perspectives

Let's return to **cilantro**, and attend to an important feature of the case that I have so far ignored. That is: although Sam can know that his evidence is misleading, he *cannot* know that *he is being misled* by his misleading evidence. A *person* is only misled by misleading evidence if she believes what that evidence supports. And Sam does not regard himself as believing what his total evidence supports. He thinks he made a mistake: that's what the newspaper article suggests. This means that when Sam decreases confidence in C, he doesn't recognize this as a departure from the truth. From Sam's *perspective* at the end of the story, C is not highly likely to be true. To return to the analogy from section 2, the position Sam *thinks* he is in is like that of someone who has deviated from the pirate's instructions, but found the treasure anyway.

More specifically, what *actually* happened to Sam was this: he accommodated his first-order evidence rationally, moving towards the truth. Then he accommodated his higher-order evidence rationally, moving away from the truth. But what Sam *rationaly thinks* happened is something else: he accommodated his first-order evidence *irrationally*, moving away from the truth. Then he accommodated his higher-order evidence rationally, moving back towards the truth. Given what Sam rationally thinks has happened to him, he should regard his low confidence in C, at the end of the story, as

³⁰ In this section I will take the phrases "what the evidence supports" and "what the evidence makes likely" to be synonymous. So, I will take (3) to be equivalent to:

(3') Rationality requires believing what your evidence supports.

We could choose instead to use talk of "evidential support" in another way, according to which one could coherently hold one of (3) and (3') and deny the other. For instance, we could instead say that "what the evidence supports" is equivalent to "what one should rationally believe given one's evidence". (Christensen [2010] writes: "HOE really is best thought of as evidence. It is information that affects what beliefs an agent (even an ideal agent) is epistemically rational in forming." (p. 193)) On that view, (3') is analytic, but not obviously equivalent to (3). For example, if a body of evidence *entails* some proposition, then it makes that proposition likely. But one might deny that it is rational to believe everything entailed by one's evidence.

I find it natural to read (3) and (3') as equivalent, but I have no particular objections to interpreting them differently. Someone who feels strongly that (3) and (3') are not equivalent should substitute talk of "evidential support" in this section with talk of "what is likely, given one's evidence".

highly accurate. But since (so he thinks) he arrived at this accurate belief by departing from what the evidence supported, he should not regard his final doxastic state as fully supported by the evidence. Building on the pirate analogy: the position Sam thinks he is in is like someone who deviated from a *mistaken* set of instructions, and found the treasure anyway.

This is not an obviously irrational attitude to have. Compare: there would be something odd about searching for treasure by precisely following a set of instructions that one takes to contain (particular) mistakes. But there would not automatically be anything odd about searching for treasure using a mistaken map *that one was not following precisely*. Similarly, it is not incoherent for Sam to believe P while taking himself to have evidence that *misleadingly* points away from P.³¹

How can we incorporate this observation into our account of higher-order evidence? We want to say that it is rational for Sam to take C to be unlikely: *from his perspective*, C is probably false. But given Sam's *total evidence*, C is still highly likely: his evidence makes C probable.

This second conclusion, that C is likely given Sam's evidence, seems hard to deny. How could one dispute it? One might try the following Very Easy Argument to argue that C is *not* highly likely given Sam's evidence:

Very Easy Argument To determine what's likely given some body of evidence, we should just think about what an agent with that total evidence *should* believe. Sam should have low confidence in C after accommodating his total evidence. Therefore, C is not highly likely, given Sam's total evidence.

Unfortunately, the Very Easy Argument is nonsense. Depending on how we set up the case (and how reliable Sam takes his roommate to be), E + HOE may in fact *entail* C.

³¹ In coming to have this attitude, Sam is "epistemically akratic": he has a belief that he believes is unsupported by his evidence. Many have taken for granted that such beliefs are irrational. Horowitz [2014] argues at length that it is irrational to have attitudes of this form – however, that paper argues, this is only true in cases where our evidence tends to be "truth-guiding". In cases where evidence is "falsity-guiding", i.e. predictably misleading, these attitudes are not so strange. If my argument here is right, higher-order evidence in general is falsity-guiding. See also Christensen [2016 a], Horowitz and Sliwa [2015], Worsnip [2018], Lasonen-Aarnio [ms], and Smithies [ms] for further discussion of epistemic akrasia in cases of higher-order evidence.

Take the strongest case, where E entails C; adding HOE to E does not destroy that entailment.³² So we should accept that C is highly likely, given Sam's evidence.

To hold both of these views about Sam's situation, we have to say that a rational agent's perspective is distinct from her total evidence; so, what she should believe is distinct from what her total evidence supports. What Sam *should believe* is what's likely from his perspective; this might or might not align with what's supported by his total evidence. If we make this distinction, we can see why it might not be so bad after all to say that higher-order evidence is predictably misleading. While one can predict *in the abstract* that higher-order evidence makes it rational to move one's beliefs farther from the truth, one cannot recognize in one's own case when this is happening.

The suggestion to distinguish between evidential support and an agent's rational perspective is not completely novel. In fact, I think this is a natural interpretation of a popular class of views about higher-order evidence, which says that rational agents should calibrate the strength of their credences to the expected reliability of their own reasoning (meaning, roughly: the extent to which that reasoning tends to point to the truth). Following recent literature, let's call the view "**Calibrationism**".³³

According to Calibrationism, the assessment of one's own reliability must be *independent* from the evidence and reasoning in question. This is because appealing to the evidence and reasoning in question is question-begging. If Sam could rationally rely on his first-order evidence while assessing his reliability, he would be able to reason like this: "The newspaper says I'm likely to reach a false conclusion about C. But this proof is valid, and my roommate said that if it was valid, C would be true. Therefore, the newspaper is wrong: I'm immune from the effects of the cilantro/soapy-taste gene!" That reasoning looks irrational – just like it would be irrational for someone to conclude that he is not colorblind just by looking at objects and noting their apparent colors. To test

³² Of course, some have argued that in some cases it is not rational to believe what is entailed by one's evidence. See, for instance, Schechter [2013]. My position here is fully compatible with this view. According to my present suggestion, we should agree with Schechter and others that it is not always *rational to believe* what is entailed by one's evidence. But it is still true that one's evidence makes its entailments *highly likely*. That is, given Sam's proof, note from his roommate, and the newspaper report, the curry *probably does* contain cilantro. See Christensen [2010] for further discussion of this point.

³³ See White [2009], Schoenfield [2015 b], Horowitz & Sliwa [2015] and Lasonen-Aarnio [ms] for further discussion; "calibrationism" is a generalization of the "conciliationist" or "equal weight" views defended by Christensen and Elga. There are different versions of the view, which I will set aside for the moment.

one's color vision, one needs *independent* confirmation that one's color-appearances are accurate. Similarly, to assess the reliability of one's reasoning in some domain, Calibrationists hold that one needs to "bracket" or "set aside" the evidence and reasoning in that domain.³⁴

I think it is natural to interpret this feature of Calibrationism as forcing a separation between one's evidence and one's rational perspective. Once evidence is "set aside", it is no longer part of one's perspective; it can't be appealed to in forming beliefs. But advocates of Calibrationist views are often careful to insist that the evidence doesn't disappear when it is "bracketed". So bracketing is a change in perspective without a change in evidence.

This means that, for the Calibrationist, the way in which higher-order evidence rationalizes belief change can't be explained by the normal notion of evidential support (again, in the sense of what the evidence makes likely). Nevertheless, higher-order evidence has a rational effect on first-order beliefs. Its effect is to change what is likely from a rational agent's *perspective*, by changing the expected reliability of the rational agent's process of reasoning.³⁵

4.2 Upshots for Calibrationism

In this final section, I will look at some potential upshots and unsolved problems for Calibrationism, given this new interpretation of the view.

An interesting feature of Calibrationism is that it seems to treat reasoning as just another source of beliefs, alongside perception, memory, and so on. This raises questions about just how far the analogy extends. For example: in cases of higher-order defeat, a rational perspective is one that disregards some evidence. Should we think of perceptual (undermining) defeat in the same way? It seems to me that there is reason to answer "yes". To avoid bootstrapping, we need to say that an agent's estimate of the reliability of her perception must be independent of the perceptual evidence in question. If that is true,

³⁴ The notion of "bracketing" is due to Elga, and has a large role in Christensen's work. For further discussion of independence, see Christensen [2009], Christensen [2007], and Elga [2007]. See Vavova [2018] for helpful discussion of how exactly these independence principles should be formulated to allow them to avoid skepticism.

³⁵ Notice that on this view, what's rational to believe still *supervenes* on one's total evidence. (We can hold that there is no change in perspective without a change in total evidence.) So Calibrationism is not necessarily incompatible with evidentialism, if the latter is understood as a supervenience claim.

it seems that perceptual defeat requires us to set aside evidence – and therefore, the phenomenon of perceptual defeat already motivates the thought that we should believe what is likely given our perspective, rather than what is likely given our evidence.³⁶

But even if there is a strong analogy between reasoning and other sources of belief, there are important disanalogies as well. One disanalogy is that while we are epistemically responsible for reasoning well on the basis of our evidence, we are arguably *not* epistemically responsible for seeing well, hearing well, remembering well, and the like. How can we maintain the thought that reasoning is epistemically assessable, if these other belief-forming mechanisms are not? One plausible option for the Calibrationist is to say that we are rationally required to reason well, *and* we are also rationally required to calibrate our beliefs to the expected reliability of their sources (including reasoning). This begins to look like a version of a two-norm view, in that there are two separate kinds of epistemic evaluation at work in cases like **cilantro**. But identifying the two modes of evaluation in this way may have better potential for answering the central challenge for two-normers, which was to explain why both norms are distinctly epistemic, and both worth caring about. Reasoning is clearly tied to the truth, in that it is the process by which we discover what follows from, or what is made probable by, our evidence. It makes sense that we care about reasoning well, from a purely epistemic point of view. But reasoning alone doesn't necessarily give an agent the best shot at truth, *from her own perspective*. Calibration does that. And it is rational to calibrate.

Finally: one might wonder, *why* is it rational to calibrate, given that *calibrating* is predictably misleading?³⁷ We can know a priori that a rational agent who calibrates on the basis of higher-order evidence, and revises her belief through that calibration, will have less accurate beliefs than an agent who does not. So why is calibrating a good idea? Here I think the Calibrationist's best defense is to say that, contra Carnap, rationality is not about programming the best robot. (*Evidential support*, on the other hand, might be.) Instead, it is about the search for truth from one's own perspective – a perspective that is, unfortunately for us, sometimes limited.

³⁶ Lasonen-Aarnio [2014] raises a similar point, extending worries about higher-order defeat to undermining defeat in general.

³⁷ Thanks to [OMITTED] for raising this worry.

5. Conclusion

I have argued that higher-order evidence is predictably misleading. Insofar as we are rational, accommodating this evidence correctly will tend to lead us away from the truth. So why should we believe what higher-order evidence supports? I have argued that one standard view of higher-order evidence may be able to answer this challenge. However, doing so comes at a cost: to maintain the standard view, we must say that we should not always believe what our evidence makes likely. This cost must be weighed against the benefits of the view: maintaining our intuitions about higher-order defeat in cases like **cilantro**, and maintaining that such defeat is distinctly epistemic phenomenon, with straightforward ties to the truth.³⁸

³⁸ [Acknowledgments]

References

- Carnap, Rudolf. [1962] "The Aim of Inductive Logic". In Sahotra Sarkar (ed.), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism*. New York: Garland. 1996. pp. 259-274.
- Carr, Jennifer Rose [2017]. "Epistemic Utility Theory and the Aim of Belief". *Philosophy and Phenomenological Research* 95 (3):511-534.
- Christensen, David. [2016 a] "Disagreement, Drugs, etc.: from Accuracy to Akrasia". *Episteme* 13 (4):397-422.
- [2016 b] "Conciliation, Uniqueness, and Rational Toxicity". *Noûs* 50 (3):584-603
- [2013] "Epistemic Modesty Defended". In *The Epistemology of Disagreement: New Essays*, David Christensen and Jennifer Lackey ed., OUP 2013.
- [2011] "Disagreement, Question-Begging and Epistemic Self-Criticism", *Philosophers' Imprint* 11, (6)
- [2010] "Higher-Order Evidence", *Philosophy and Phenomenological Research* 81, 185-215
- [2009] "Disagreement as Evidence: The Epistemology of Controversy". *Philosophy Compass* 4 (2009), 756-67
- [2007] "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals", *Oxford Studies in Epistemology* 2, 3-31
- Elga, Adam. [2007] "Reflection and Disagreement". *Noûs* 41 (3): 478-502.
- Feldman, Richard. [2007] "Respecting the Evidence". *Philosophical Perspectives volume 19: Epistemology*. (Blackwell, Oxford) pp. 95-119.
- Horowitz, Sophie. [2014] "Epistemic Akrasia". *Noûs* 48 (4):718-744.
- Horowitz, Sophie and Paulina Sliwa. [2015] "Respecting All the Evidence." *Philosophical Studies* 172 (11):2835-2858.
- Kelly, Thomas. [2010] "Peer Disagreement and Higher Order Evidence." in Richard Feldman and Ted Warfield (eds.), *Disagreement* (OUP): 111-174.
- Konek, Jason and Ben Levinstein, "The Foundations of Epistemic Utility Theory". Forthcoming in *Mind*.

- Kvanvig, Jonathan. [2014] *Rationality and Reflection*. OUP.
- Lasonen-Aarnio, Maria. “Enkrasia or Evidentialism”. Forthcoming in *Philosophical Studies*.
- [2014] “Higher-Order Evidence and the Limits of Defeat.” *Philosophy and Phenomenological Research*. 88.2: 314-345.
- Schoenfield, Miriam. [2015 a] “Bridging Rationality and Accuracy.” *Journal of Philosophy* 112 (12):633-657.
- [2015 b] “A Dilemma for Calibrationism”. *Philosophy and Phenomenological Research* 91(2): 425-55.
- Pryor, James. [2004] “What’s Wrong With Moore’s Argument?” *Philosophical Issues* 14: 349-378.
- Schechter, Joshua. [2013] “Rational Self-Doubt and the Failure of Closure.” *Philosophical Studies* 163 (2): 28–452.
- Sepielli, Andrew. [ms] “Evidence, Reasonableness, and Disagreement.”
- Smithies, Declan. [ms] “The Irrationality of Epistemic Akrasia.”
- Vavova, Katia. [2018] “Irrelevant Influences”. *Philosophy and Phenomenological Research* 96 (1):134-152.
- Weatherson, Brian. [2014] “Running Risks Morally”. *Philosophical Studies* 167: 141-163.
- White, Roger. [2009] “On Treating Oneself and Others as Thermometers.” *Episteme*.
- Williamson, Timothy. [2000] *Knowledge and its Limits*. OUP.
- Worsnip, Alex. [2018] “The conflict of evidence and coherence.” *Philosophy and Phenomenological Research* 96 (1):3-44.