

# Higher-Order Defeat and Doxastic Resilience

Asbjørn Steglich-Petersen

Penultimate draft. Final version to appear in M. Skipper & A. Steglich-Petersen, *Higher-Order Evidence: New Essays*. Oxford University Press

**Abstract:** It seems obvious that when higher-order evidence makes it rational for one to doubt that one's own belief on some matter is rational, this can undermine the rationality of that belief. This is known as higher-order defeat. However, despite its intuitive plausibility, it has proved puzzling how higher-order defeat works, exactly. To highlight two prominent sources of puzzlement, higher-order defeat seems to defy being understood in terms of conditionalization; and higher-order defeat can sometimes place agents in what seem like epistemic dilemmas. This chapter draws attention to an overlooked aspect of higher-order defeat, namely that it can undermine the *resilience* of one's beliefs. The notion of resilience was originally devised to understand how one should reflect the 'weight' of one's evidence. But it can also be applied to understand how one should reflect one's higher-order evidence. The idea is particularly useful for understanding cases where one's higher-order evidence indicates that one has failed in correctly assessing the evidence, without indicating whether one has over- or underestimated the degree of evidential support for a proposition. But it is exactly in such cases that the puzzles of higher-order defeat seem most compelling.

## 1. Introduction

It seems obvious that when higher-order evidence makes it rational for one to doubt that one's own belief or credence on some matter is rational, this in itself can undermine the rationality of that belief or credence. For example, evidence that I suffer from the common self-enhancement bias undermines the rationality of my belief that I am a better than average cook. This phenomenon is known as higher-order defeat. However, despite its intuitive plausibility, it has proved puzzling how higher-order defeat works, exactly. To highlight two prominent sources of puzzlement, higher-order defeat seems to defy being understood in terms of conditionalization, since higher-order evidence does not affect the probability of the contents of the beliefs it undermines. And higher-order defeat can place agents in what seem like epistemic dilemmas, when the first- and higher-order evidence pull in different directions.

In this chapter, I try to make progress on these issues by drawing attention to an overlooked aspect of higher-order defeat, namely that it can undermine the *resilience* of one's beliefs. This aspect has been noted briefly by Andy Egan and Adam Elga (2005), but has not yet received systematic treatment.<sup>1</sup> The notion of resilience was originally devised to understand how one should reflect the 'weight' of one's evidence in one's beliefs. But I argue that it can also be applied to understand how one should reflect one's higher-order evidence. The idea is particularly useful for understanding cases where one's higher-order evidence indicates that one has failed in correctly assessing the evidence, without indicating in what direction one has erred, i.e. whether one has over- or underestimated the degree of support for a proposition from one's evidence. But as I shall argue, it is exactly in such cases that the puzzles of higher-order defeat seem most compelling.

---

<sup>1</sup> Roger White (2009) and Alex Worsnip (2014) both discuss the significance of resilience for how to respond to epistemic disagreement. However, while they focus on the resilience of one's estimate of one's own (and one's interlocutor's) reliability, and the significance of this for epistemic disagreement, I am interested in the impact of higher-order evidence on the resilience of one's first-order credences.

Here is how I will proceed. In §2, I motivate that in some cases, it can be rational to maintain one's level of credence despite rational doubt that the credence is rational. In §3, I argue that in such cases, one should respond by lowering the resilience of one's credence, rather than its level, and sketch how the relevant changes in resilience can be understood in terms of changes in an agent's higher-order credences, which are directly constrained by one's higher-order evidence. In §4, I discuss why higher-order doubt undermines credences and categorical beliefs in importantly different ways, and how the idea of resilience-defeat can explain this. In §5, I show how the proposed account can help resolve the puzzles concerning conditionalization and epistemic dilemmas. §6 is a brief summary.

## 2. Rational Credence in the Face of Rational Doubt

I will begin by considering some cases where it seems rationally permissible for one to maintain a certain level of credence in a proposition, even in light of rational doubt that this is the right level of credence to hold in one's epistemic situation. Consider the following case, described by David Christensen (2010a: 121):

**Doubtful Ava:** “[...] Ava is considering the possibility that (D) the next U.S. President will be a Democrat. She gives D some particular credence, say .7; this reflects a great deal of her general knowledge, her feel for public opinion, her knowledge of possible candidates, etc. But given the possibility that her credence is affected by wishful thinking, protective pessimism, or just failure to focus on and perfectly integrate an unruly mass of evidence, Ava very much doubts that her credence is exactly what her evidence supports.”

As Christensen notes, it seems ‘eminently reasonable’ for Ava to doubt that her level of credence is exactly right, but this in itself doesn't seem to undermine the rationality of her maintaining that level of credence. That is not to say, of course, that Ava's credence could not be irrational for other reasons. Perhaps she really did misinterpret her evidence, thus making her credence irrational in light of her first-order evidence alone. But the fact that she harbours rational doubt that her credence is rational in light of the evidence, should not in itself move her to revise it.

Similar examples are easy to come by. Consider the following case, inspired by Kahneman and Tverski's famous discussion of the base rate fallacy (1982):

**Doubtful Bob:** A taxicab has been involved in a hit and run accident at night, and Bob has been called to serve on a jury in court, where a driver from Blue Cabs stand accused. There are just two taxi companies in town, Green Cabs being the dominant one. Only 5% of the taxis in town are blue, 95% are green. A witness saw the incident, and identified the cab as blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time, and failed 20% of the time. Bob is carefully considering how likely it is that the cab was blue rather than green in light of the available evidence, and ends up giving it credence .17. However, Bob always found probabilities tricky, and his grasp of Bayes' Rule is pretty hazy. He thus severely doubts that his credence is right given the evidence. In fact, he would not be very surprised if he is quite far off the mark.

Again, given Bob's insight into his own imperfect understanding of probabilities, it seems reasonable for Bob to doubt the correctness of his credence. But again, this does not seem to by itself make it irrational for Bob to maintain this credence. He has considered his evidence carefully, applied principles of probabilistic reasoning to the best of his ability, and can't make the result come out any other way. So if his doubt is to have an impact on his credence, it is unclear in what direction he should revise it. It therefore doesn't seem as if his doubt by itself should make Bob adopt a different level of credence.

Not all cases of rational higher-order doubt allow one to maintain one's credence. Indeed, the kind of cases that epistemologists have tended to focus on, do not seem to allow this. Immediately after presenting the case I've called Doubtful Ava, Christensen goes on to present a case of higher-order doubt that clearly does require a revision of credence (2010a: 121). Here, Brayden, a staunch Republican, also initially gives credence .7 to D, but then receives compelling evidence of his own tendency to become irrationally confident of unpleasant possibilities, and to never underestimate them. This makes it rational for Brayden to believe that his credence is irrationally high, which plausibly means that he should lower his credence in D. The difference between this case and Doubtful Ava, Christensen notes, is that while Ava's higher-order evidence of her own fallibility does not 'lopsidedly' suggest that her credence deviates from what is rational in one particular direction, Brayden's does. Most cases of higher-order defeat discussed in the literature are lopsided in this way. Cases of revealed peer-disagreement, for example, where two peers discover that they differ in the level of credence they have adopted in response to the same evidence, seem to have this property.

By contrast, I want to focus, at least initially, on the 'non-lopsided' cases, like Doubtful Ava and Doubtful Bob. As will become clear, I think that understanding the kind of defeat involved in these cases is helpful for understanding higher-order defeat in general. The plausibility of such cases does not depend on exactly how much doubt in the correctness of their own credence it is rational for Ava and Bob to have. The cases are not very specific in this regard, and it seems that a relatively generous range of doubt, from slight to quite severe, leaves the intuition that it doesn't require a revision of credence intact. Ava 'very much doubts' that her credence is right, yet it is rationally permissible for her to maintain it. The same is the case for Bob. This leaves it possible there is an upper limit to how much doubt in one's own credence rationality allows. I will return to that question later. But rationality allows at least a relatively high degree of doubt.

However, even if Ava and Bob aren't rationally required to revise their level of credence, it seems that their doubt ought to be reflected *somehow* in their credences. But how? The proposal that I want to explore in the following is that while Ava and Bob may maintain their *level* of credence, their credences should become less *resilient* as a result of their rational doubt. I now turn to introduce and motivate that idea.

### 3. Doxastic Resilience and Higher-Order Doubt

The notion of resilience was originally introduced to explain how evidential *weight* should be reflected in one's credences. Keynes introduces the notion of evidential weight as follows:

"As the relevant evidence [for a hypothesis] at our disposal increases, the magnitude of [its] probability may either decrease or increase, according as the new knowledge strengthens the unfavourable or favourable evidence; but *something* seems to have increased in either case – we have a more substantial basis on which to rest our conclusion. [...] New evidence will sometimes decrease the probability of [the hypothesis] but will always increase its 'weight'." (1921: 77).

To illustrate this idea, suppose that you have found a coin left behind by a deceased cardsharp.<sup>2</sup> You have some reason to suspect that the coin is biased, but you are not sure of this, and have no idea to what side it is biased, if it is. You can now imagine tossing the coin a number of times, and consider what credence you should adopt that the 100<sup>th</sup> toss will yield heads, given various series of prior results. Before tossing the coin at all, you have very little information, but the little you have points equally strongly towards heads and tails on the 100<sup>th</sup> toss. It thus seems that in that situation, you should adopt credence .5 in heads. Suppose now that you have tossed the coin 90 times, and that the results are divided evenly between heads and tails. Still, it seems that you should have a credence of .5 in the 100<sup>th</sup> toss landing heads. However, something has obviously changed: you now have a much weightier basis for your credence.

How should this increased weight be reflected in your credence, if not in its level? One plausible answer is that it should be reflected in what Brian Skyrms (1977) and others have called its degree of ‘resilience’, which can be understood as measuring how much the level of credence should change in the face of additional data. Imagine, for example, that we add a series of five heads to the two evidential situations above. Adding this series to the situation where you have you haven’t yet observed a single toss should clearly have a large impact on your credence that the 100<sup>th</sup> toss will land heads. You should move from .5 to something much closer to 1. But adding it to the situation where you have already observed 90 tosses should make a relatively modest impact. Your level of credence in this latter situation is thus much more resilient than in the first.<sup>3</sup>

Although the idea of resilience was originally introduced to explain how one should reflect the weight of one’s evidence, it seems that other factors besides weight should have a similar effect. Egan and Elga (2005: 77-79) observe that when one’s credence is based on how reliable one takes some channel of information to be, changes in this reliability should be reflected not only in the level of one’s credence, but also in its resilience. Suppose that you seem to remember that the person across the room is called Sarah. If you rationally regard your memory for names as 99% reliable, you should adopt a credence of .99 that the person’s name is indeed Sarah. If it becomes rational for you to regard your memory as less reliable, you should adjust your credence in its deliverances accordingly. For example, were you to become rationally convinced that your memory is only 90% reliable, your credence should be .9, etc. But this reduction of trust in your own memory should have a further effect: it should make your memory-based credences less resilient. To show this, Egan and Elga consider what would happen if you overheard someone else calling the person across the room ‘Kate’ instead of ‘Sarah’, and you rationally regard this person as 95% reliable in remembering names. In the situation where you regard yourself as 99% reliable, this experience should make you adjust your credence only moderately, to .84. But in the situation where you regard yourself as 90% reliable, the reduction you should make is rather drastic, namely to .32.<sup>4</sup> So reduced trust in your channel of information should affect both your level of credence, and the sensitivity of this credence to new information, i.e. its resilience.

This indicates that resilience is a property that should be affected by several different factors, evidential weight and reliability of one’s informational source being two examples. I now want to argue that changes in rational doubt about the correctness of one’s level of credence in light of the evidence, should have a similar effect on the resilience of one’s

---

<sup>2</sup> This example originates in Popper’s discussion of the paradox of ideal evidence (1972); this particular version is inspired by one found in Hansson (2008).

<sup>3</sup> For a helpful discussion of the impact of evidential weight on doxastic resilience, see Joyce (2005).

<sup>4</sup> As they note, this assumes that independently of your memory impressions, you regard the two names as being equally likely to be correct (2005: 78).

credence. To give some initial motivation for this idea, consider the following variation of Doubtful Ava:

**Confident Eve:** Eve is considering the possibility that (D) the next U.S. President will be a Democrat. She gives D some particular credence, say .7; this reflects a great deal of her general knowledge, her feel for public opinion, her knowledge of possible candidates, etc. As Eve is well aware, she is even-minded and very experienced in integrating unruly bodies of evidence, and she is therefore rationally highly confident that her level of credence reflects what her evidence supports.

While it seems rationally permissible for both Ava and Eve to maintain credence .7 in D, it seems plausible that Ava's credence should be less resilient than Eve's in the light of new evidence that they might receive. Suppose, for example, that Ava and Eve both receive some new evidence which clearly speaks against D, but not decisively so, namely that a heightened threat of terrorist attacks has moved some voters to prefer a more hawkish national security policy, traditionally associated with the Republicans. For simplicity we can assume that both Ava and Eve correctly understand the significance of this new evidence, and that they are both rationally confident of what this is. How should their level of credence in D change once they add the new evidence?

Here it seems plausible that Doubtful Ava should move to a lower credence than Confident Eve. Ava, after all, was unsure, and was reasonable to be unsure, that she was correct in taking the original evidence to speak in favour of D as much as her credence reflected. So when she becomes rationally convinced that the new evidence speaks against D, this should pull her a fair bit in the other direction. Eve, on the other hand, was rationally convinced that she was correct in taking the old evidence to support her credence, so she can place more weight on the old evidence when taking the new evidence into account.

A similar verdict seems plausible if we compare Doubtful Bob to a more confident juror, Bill:

**Confident Bill:** A taxicab has been involved in a hit and run accident at night, and Bill has been called to serve on a jury in court, where a driver from Blue Cabs stand accused. There are just two taxi companies in town, Green Cabs being the dominant one. Only 5% of the taxis in town are blue, 95% are green. A witness saw the incident, and identified the cab as blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time, and failed 20% of the time. Bill is carefully considering how likely it is that the cab was blue rather than green in light of the available evidence, and ends up giving it credence .17. Bill always loved probability puzzles, and is a long time student of Bayesian reasoning. He is thus rationally very confident that his credence is right given the evidence.

Again, although their different levels of rational doubt allow both Bob and Bill to maintain the same level of credence, it seems plausible that their credence should differ in sensitivity to new evidence. Suppose that the prosecutor introduces a new piece of evidence: a blue paint trace from the crime scene. The paint is of a common kind, and it cannot be dated with accuracy, so the new evidence is not conclusive. But it certainly does support to some degree that the guilty cab was blue. Suppose that both Bob and Bill become rationally convinced of the significance of this new evidence. How should they react? It seems plausible that Bob

should revise his credence more than Bill should. Prior to receiving the new evidence, they both had a relatively low credence that the cab was blue, but only Bill was rationally confident that this low credence adequately reflected his evidence. Bill should thus place more weight on this prior evidence, than Bob should, and thus be moved less in the direction suggested by the new evidence.

In these examples, the difference in resilience showed up in how much the relevant credences should be affected by new evidence bearing directly on the object-level proposition, i.e. that the next president will be a Democrat, and that the cab at the crime scene was blue. But it seems that there should be a similar difference in resilience in the face of new higher-order evidence, for example in the form of revealed disagreement. Suppose, for example, that Ava and Eve both meet another person, Adam, that after careful consideration of their evidence judges it to support a credence in  $D$  of .3. Insofar as Ava and Eve place at least some confidence in Adam's ability to assess the evidence, they should both reduce their level of credence to some degree. But again, it seems that Ava should move her credence more than Eve in the direction of Adam's. So the lower resilience of Ava's credence should manifest itself in sensitivity to new evidence of both first- and higher-order (more on such cases below).

Obviously, if differences in how rationally confident we should be in our own assessments of the evidence make a difference to how resilient our credences ought to be, higher-order evidence bearing directly on how confident we should be in our own assessments will normatively constrain the resilience of our credences. If Doubtful Ava acquires compelling evidence that she is actually less prone to bias and wishful thinking than she thought, she should become more confident in her assessment of the evidence, and her credence should thus become more resilient. And if Confident Eve acquires compelling evidence that she is actually more prone to bias and wishful thinking than she thought, she should lower her confidence in her own assessment of the evidence, and her credence should thus become correspondingly less resilient. If we think of higher-order evidence as evidence that concerns how rationally one has responded to one's first-order evidence, this suggests a new normative role for higher-order evidence, namely as something that should affect the resilience of one's credences. More specifically, our higher-order evidence constrains how confident we ought to be that our first-order credences correctly reflect our evidence. It constrains, in other words, one's 'higher-order credences' about the correctness of one's first-order credences. And how confident we ought to be that our first-order credences correctly reflect our evidence in turn constrains how resilient those first-order credences ought to be.

How exactly do rational higher-order credences constrain the resilience of the first-order credences they take as their object? This can be seen by reflecting further on cases of epistemic disagreement.<sup>5</sup> As mentioned above, the resilience of a first-order credence that  $p$  can be a matter of sensitivity to both first-order evidence bearing directly on  $p$ , and higher-order evidence indicating that a level of credence different from one's own better reflects the first-order evidence.

Begin with the latter kind, i.e. sensitivity of a first-order credence in  $p$  to new higher-order evidence concerning the correctness of that credence. This sort of case is familiar from the literature on epistemic disagreement. Suppose that you have adopted some particular level of credence in response to some evidence, and then meet someone who has adopted a different level of credence in response to the same evidence. You thus disagree about what

---

<sup>5</sup> I assume here a broadly conciliatory approach to disagreement, such as that defended by e.g. Christensen (2007). For a precise probabilistic account of how peer-disagreement can be understood in terms of higher-order evidence, see Rasmussen et al. (2018).

credence your shared evidence supports. How should you react? That depends on how reliable it is rational for you to regard yourself and your friend as in assessing the evidence, i.e. how likely you are rational to think it is for each of you to have assessed the evidence correctly. Other things being equal, the more likely it is for you that you have evaluated the evidence correctly, the less you should revise your credence in the direction of your friend's, and *vice versa*. So rational higher-order credences about how well your first-order credences reflect your evidence affect how sensitive those credences should be to new higher-order evidence, i.e. their resilience in the face of such evidence.

But how can sensitivity to new first-order evidence bearing directly on the object-level proposition be understood in terms of higher-order credences? Again, cases of disagreement provide a useful model, although this time it requires taking a step back. Philosophers have tended to focus on cases where the disagreeing parties have responded differently to *the same* shared body of evidence, but it seems that parties to a disagreement should respond in more or less the same way if their evidence isn't identical, but merely *on par*. Christensen (2007) considers a case where I have good reason to believe that my friend's evidence with respect to *p*, although different from mine, is just as good. But whereas his evidence supports his credence in *p* of .2, my evidence supports my credence of .8. For simplicity, let us assume that the two bodies of evidence are on par with respect to evidential weight. The effect on resilience generated by weight should thus be the same for both bodies of evidence. As Christensen observes, it seems that even if we have responded to different bodies of evidence, we should both revise our credences in the direction of the other.

Furthermore, in such cases, changes in cognitive parity should affect what the disagreeing parties should do, in the same way as in cases where the parties have the same evidence. If I discover that I am in fact my peer's cognitive superior, I should revise less in his direction, and *vice versa*, even if we have responded to different bodies of evidence. In fact, this seems to hold even when the evidence possessed by the disagreeing parties is not only non-identical, but also not on a par. When my evidence with respect to *p* is better than my disagreeing friend's evidence, I have some reason in virtue of that to favour my own level of credence in *p* over my friend's. But I still cannot completely disregard my friend's assessment of his evidence, and should thus revise in his direction, albeit less so than if his evidence had been as good as mine. Conversely, when my friend's evidence is better than mine, I have reason to favour his level of credence, but needn't completely disregard my own. And again, even in situations such as these, changes in cognitive parity should matter to how much I should revise. Even in cases where my evidence is better or worse than my friend's, I should revise less if I discover that I am his cognitive superior, and more if I discover that he is cognitively superior to me.

With this in mind, return now to the question of sensitivity to new first-order evidence. Suppose that I have adopted credence .8 in *p* in response to one body of evidence *E*<sub>1</sub>, and that I rationally regard myself as highly reliable in evaluating *E*<sub>1</sub>. I now get some new evidence bearing on *p*, *E*<sub>2</sub>, which is on a par with *E*<sub>1</sub> in terms of evidential weight. Again, I rationally regard myself as highly reliable in evaluating *E*<sub>2</sub>, and judge that this evidence, in isolation, supports a credence of .2 in *p*. What credence in *p* should I adopt in light of my total evidence? The conflicting bodies of evidence are on a par, and I rationally regard myself as equally reliable in evaluating both. So it seems that by reasoning parallel to that used in cases of epistemic disagreement, I should adopt a credence roughly in the middle, i.e. .5.<sup>6</sup>

---

<sup>6</sup> This assumes, of course, that the two bodies of evidence are independent, and that there are no undercutting defeating relations between them.

But what if I rationally regard myself as being less reliable in evaluating E1 than E2, perhaps because of the mathematics involved in that body of evidence, and thus rationally regard it as less likely that I have evaluated E1 correctly than E2? I would then seem to find myself in a situation parallel to the case of disagreement with evidential parity but cognitive non-parity. In that case, I should thus adopt a credence closer to that which I judge to be supported by E2 alone, than that which I judge to be supported by E1 alone. On the other hand, if I rationally regard myself as more reliable in evaluating E1 than E2, I should end up with a credence closer to that which I judge to be supported by E1 alone. Again, as in the case of disagreement, this can be generalised beyond cases with parity between conflicting bodies of evidence. Regardless of whether E1 and E2 are on a par as evidential bodies, *ceteris paribus* I should end up with a credence closer to, or farther from, that which I judge to be supported by E1 alone, depending on how reliable I rationally think I am at correctly assessing the import of E1.

So it seems that cases of epistemic disagreement with non-identical evidence provides a model for how to understand the resilience of first-order credences in the face of new object-level evidence, in a way that relies on rational higher-order credences about one's reliability in assessing the respective bodies of evidence. The resilience of a first-order credence in the face of new first-order evidence is affected by how reliable one regards oneself as being in assessing the new and old evidence, in roughly the same way as the sensitivity of one's credence in cases of disagreement depends on how reliable one thinks that oneself and the disagreeing friend are.

The resulting picture is that higher-order evidence of one's own rational failure undermines the resilience of one's first-order credences, by undermining how likely it is rational for one to think that those credences correctly reflect the evidence on which they are based. This in turn makes those credences more sensitive to both higher-order evidence that a different credence level is correct, as in cases of disagreement, and to new first-order evidence supporting a different credence level; in other words, it makes those credences less resilient.

So far, I have focused on cases of higher-order defeat which allows one to maintain one's level of credence. I have argued that in such cases, the defeat should be understood as undermining the resilience of one's credence, instead of its level. But not all cases of higher-order defeat are like that. In some cases, the evidence not only lowers the probability that one's credence correctly reflects the evidence, but also 'lopsidedly' indicates in what direction one has erred. This is the case in Christensen's example of Brayden the Republican, and in cases of epistemic disagreement.

Both kinds of higher-order defeat can be understood in terms of the effect they have on one's higher-order credences about what first-order credence best reflects the evidence. In lopsided cases, the higher-order evidence makes it more likely for one that another credence is correct in light of one's evidence, which means that one must revise one's level of credence. How much one should revise depends on how probable it is for one that one's original level of credence was correct, as well as how probable the higher-order evidence makes it that another level is correct.<sup>7</sup> In non-lopsided cases, defeating higher-order evidence makes it less likely for one that one's credence is right in light of one's first-order evidence, but does not make another level of credence more likely to be correct than the original one. It 'spreads out' or 'flattens' one's higher-order probability distribution, without pulling its highest point in a particular direction. This means that what's defeated is not the level of one's first-order credence, but only its resilience. In general, then, higher-order defeat works

---

<sup>7</sup> For a detailed account of this, see Rasmussen et al. (2018).



by affecting one's higher-order probabilities, but only in some cases does this require a change in credence.

#### 4. Resilience Defeat and Categorical Belief

In my discussion so far, I have focused on the effects of higher-order doubt and evidence on credences or graded beliefs. One might reasonably wonder how well the above account applies to categorical beliefs. On the face of it, it may seem as if rather little applies. Consider the kind of cases where it seemed plausible that some particular level of credence in a proposition could be rationally combined with a fairly high degree of doubt that this level of credence was correct in light of the evidence. Could similar cases be thought of involving categorical beliefs? That is, could it be rational to believe that *p* outright while harbouring a high degree of rational doubt that it is rational or justified to believe that *p* in light of one's evidence? That seems much less plausible. It may well be possible to rationally believe that *p* while having some relatively low degree of rational doubt about this belief, but the room for this seems much more limited. This difference ought to strike us as puzzling. Both credences and categorical beliefs must be based on adequate evidence in order to be rational or justified. So why can it be rationally permissible to maintain a credence in the face of rational doubt about its rationality, when this isn't permissible with categorical beliefs?

In addition to being a puzzling difference in itself, something that we ought to try to explain, the difference might also be seen to undermine the picture that I have proposed of higher-order defeat. This is because there is usually thought to be a tight connection between one's beliefs and credences. On one popular account of the relation, categorical belief *simply is* a suitably high degree of credence.<sup>8</sup> The picture suggested above allows one to maintain one's credences in the face of rational doubt about those very credences. So if categorical belief is simply a high degree of credence, my proposal would implausibly allow categorical belief in the face of higher-order doubts too.<sup>9</sup>

Fortunately, there is a plausible alternative theory of the relation between credence and categorical belief that allows my account to both explain the puzzling difference, and defuse the challenge it threatens to raise. This is the 'stability theory' of belief, developed in recent writings by Hannes Leitgeb (2014; 2015; 2017), and attributed by him and others to David Hume, as an early proponent.<sup>10</sup> According to the stability theory, a high credence in *p* is necessary but not sufficient for belief that *p*. In order to count as a belief, the high credence must also be suitably 'stable,' which enables belief to "play its characteristic functional role in decision-making, reasoning, and asserting [...] in the course of processes such as perception, supposition and communication" (Leitgeb 2015: 146). Leitgeb understands the stability of a credence as a matter of the credence being preserved under conditionalization, that is, under the supposition of new evidence or information (2015: 156). In other words, stability is a matter of resilience, in roughly the same sense as I have used that term above.

---

<sup>8</sup> This is what Richard Foley identifies as the 'Lockean Thesis' (1993, Ch.4). For further discussion of this and other accounts of the relation between credence and categorical belief, see Keith Frankish (2009).

<sup>9</sup> The tension between the present theory of higher-order defeat and the Lockean thesis may be less severe than this suggests, for two reasons. First, if categorical belief requires a high degree of credence, it may be hard to find realistic cases where resilience is lowered dramatically while leaving the credence level intact or above the level required for belief. This is because rational doubt that a high credence level is correct leaves more room for error in the 'too high' direction than in the 'too low' direction. Second, in some cases where it is rational for to doubt that one's credence level is correct, the margin of error may be too small to reach below the threshold for belief. I am grateful to David Christensen for pointing this out to me.

<sup>10</sup> For this interpretation of Hume's account of belief, see in particular Louis Loeb (2002). For Hume's own presentation, see his (1739-40: Bk. I, Pt. III, § 7).

I will not specify or motivate the stability theory of belief further here.<sup>11</sup> What interests me is how the theory combines with the above theory of higher-order defeat. The difference between high credence and categorical belief implies that the two states are subject to different rationality constraints. While it can be rational to have a high credence without it being rational for this credence to be very resilient, the rationality of categorical belief requires the rationality of both a high credence and a high degree of resilience. This is what explains why a high credence in a proposition is rationally compatible with a high degree of rational doubt in that credence being correct, while this is not possible for categorical belief. A high degree of rational doubt about the correctness of one's own credence would exclude the rationality of the credence being sufficiently stable, and would thereby exclude the rationality of what is necessary for outright belief, namely stable high credence. This also defuses the challenge raised by the puzzling difference between credences and beliefs, for my account of higher-order defeat. Indeed, if rational belief requires rational stability, while rational high credence does not, and higher-order evidence sometimes undermines rational stability without undermining a high degree of credence, we should *expect* this apparently puzzling difference. Rather than undermining it, the difference between credences and beliefs in sensitivity to higher-order doubt thus becomes a point in favour of the proposed account.

## 5. Applications

Thinking of higher-order evidence and defeat as something that undermines the resilience of one's first-order credences can help explain a number of puzzling features of higher-order evidence and defeat. Here, I want to focus on two such puzzling features, concerning (i) belief revision by conditionalization, and (ii) how one should respond to misleading higher-order evidence.

### 5.1 Higher-Order Evidence and Conditionalization

The first problematic feature of higher-order evidence is, in short, that it seems to give rise to mismatches between the level of confidence one should adopt in a proposition on the *supposition* that one acquires certain evidence, and the confidence one should adopt in the proposition if one *actually* acquires that evidence. Consider the following example adapted from Christensen (2010b).

A scientist is contemplating the confirmation that some currently unrealised experimental result E would provide for a hypothesis H, and judges correctly that it would provide a high degree of confirmation. He thus regards it as highly probable that H is true on the supposition that E. This makes him judge that if he actually learns that E, he should become highly confident that H is true. However, the scientist also considers how probable H would be on the supposition of E along with a further factor D: that he has been given a powerful reason-distorting drug prior to assessing his experimental results. Reasonably, the scientist realises that this addition should have no effect whatsoever on the probability of H. E confirms H to the same degree, whether or not he, or anyone else for that matter, happens to be unable to rationally assess this. So he regards it as highly probable that H is true on the supposition of E&D. But here's the catch: even though the scientist is convinced of this conditional probability, he should not resolve to become highly confident that H is true if he actually learns E&D. In fact, if he were to learn E&D, he should not be very confident

---

<sup>11</sup> I see no reason to suppose that the stability theory is incompatible with the teleological account of belief that I have defended elsewhere (e.g. 2006; 2009).

of H at all. So we now appear to have a puzzling mismatch between the scientist's prior conditional probability, and how confident he should become upon acquiring the evidence, thus contradicting the standard updating model.

By understanding higher-order defeat as something that, at least in non-lopsided cases, undermines the resilience of a credence, rather than its level, we can uphold the standard updating model. Suppose that the scientist assesses the prior probability of H given E to be .9. If he were to learn E, he should thus adopt credence .9 in H. As we have supposed, the scientist ascribes an identical probability to H given E&D. He should thus also adopt credence .9 in H if he were to learn E&D. But if he were to learn E&D, he should become much less confident that .9 is the right level of credence. As we have already seen, this is not necessarily irrational. Having a particular level of credence can be rationally combined with a high degree of doubt in that level being correct. This, however, should make the credence less resilient, i.e. more susceptible to being revised in the face of new evidence. The higher-order evidence is thus allowed to make its mark on the scientist's credence without giving up conditionalization.

If we understand the reduced confidence in one's credence being correct in terms of higher-order credences, we can understand that too in terms of conditionalization. But what is the relevant prior conditional probability? As we have already seen, D does not affect the probability of H. Does D somehow affect the prior conditional probability of H given E? That is, does the fact that the scientist is drugged somehow affect how probable the hypothesis is in light of the experimental result, for example by making it less likely that that conditional probability is of a certain level? Again, the answer is clearly no. But something that D clearly does affect, is the probability of the scientist correctly assessing the support for H given by E. More specifically, the scientist should regard it as less probable that the assessment he reaches of H given E is correct on the supposition of D, compared to the probability of that on the supposition of not-D.

So now we can explain both the credence in H that the scientist should adopt, and the low resilience that credence should have, in terms of conditionalization. Consider first the scientist's credence towards H. Prior to receiving the experimental result E, he holds the probability of H given E to be .9. He also holds the probability of H given E&D to be .9. So when he learns E&D he should adopt credence .9 in H. So how does learning D impact his credence? Call the proposition that the scientist will correctly assess the support for H given by E 'C'. Prior to learning D, the scientist ascribes some relatively low probability to C given D, say .3. So, when the scientist learns that D, he rationally adopts credence .3 in C, which, given his actual assessment of H given E, amounts to having a credence of .3 that his credence of .9 in H is correct. As long as he regards a credence of .9 in H as being more likely to be correct than any other level of credence, this can be a rational combination of credences. It just means that his credence in H should not be very resilient. The result is that the scientist has revised by conditionalization in response to both E and D, without letting D affect the probability of H, while allowing D to make its mark on the credence in H by reducing its resilience.

Here's a possible objection. Above I characterised the problem as that of explaining the mismatch between how probable one should regard a proposition on the supposition of some evidence, and the 'level of confidence' one should adopt in the proposition if one actually acquires that evidence. But does the above really explain that? After all, if we by 'level of confidence' mean level of credence, my proposal does not deliver this result. Indeed, on my proposal, one's posterior credence should match the prior conditional probability. However, I think that the intuition that one should become less 'confident' upon acquiring the higher-order evidence trades on an ambiguity in that notion. Having a high credence in

a proposition is the typical understanding of what it means to be highly confident in that proposition. But if one is highly unsure about one's level of credence being right, thus making the credence easy to affect by new evidence, it seems wrong to characterise having such a credence in a proposition as being 'highly confident' that the proposition is true, even if the level of credence is high. If so, my proposal does deliver the correct verdict.

What about lopsided cases? If the challenge to the standard updating model persists for such cases, the above solution might not be much of a solution at all. For example, suppose that the reason-distorting drug is designed specifically to dispose people to overestimate the evidence and adopt a higher credence than the evidence actually warrants. Surely, learning that one has been given such a drug should lead one to reduce one's level of credence. Yet, shouldn't one's prior conditional probability of H given E&D still match that of H given E alone? After all, whether one has been drugged has no bearing on the actual evidential bearing on E for H. If so, the mismatch between the prior conditional probability and the credence one should adopt upon acquiring the evidence seems to persist.

However, it is not obvious to me that one's prior conditional probability of H given E&D should be the same as that of H given E in lopsided cases. Consider the following lopsided case, where this seems particularly dubious. Instead of considering the probability of H given the experiment and being drugged, the scientist now considers the probability of H given the experiment and being told by his colleague Stephen Hawking that the probability of H given E is .1, and not .9 as the scientist is independently inclined to judge. If learning this, the scientist should clearly adopt a credence a fair bit below .9. But it seems equally clear that his prior conditional probability given this should be below .9 as well, thus retaining the match between the two probabilities. The scientist should clearly not be very confident that if he were to learn of the experiment and of Stephen Hawking's testimony, then H would be true, despite his actual confidence in the support for H provided by E. But why think, then, that above case with the overestimation-drug is any different? After all, Stephen Hawking's testimony is also evidence that the scientist's initial judgment overestimates the evidential import of the experiment.

This suggests that what's driving the intuition in Christensen's original case of the reason distorting drug is that in that case, the drug is not stipulated to have a lopsided effect. It is this fact that makes it plausible that having taken the drug doesn't affect the conditional probability of H given E, while at the same time requiring a lowered confidence upon learning that one has taken it. But as argued above, this lowered confidence can be understood in terms of resilience, rather than credence level. If this is plausible, neither lopsided nor non-lopsided cases exhibit mismatches between the prior conditional probability, and the level of credence one should adopt upon acquiring the evidence.

### ***5.2 Misleading Higher-Order Evidence and Akrasia***

Another major source of interest in the nature of higher-order evidence stems from cases where it appears to undermine the otherwise highly plausible idea that it cannot be rational to hold a belief that one believes to be irrational, i.e. that epistemic akrasia is always irrational. Despite the intuitive appeal of this, a number of authors have argued that in situations with sufficiently compelling misleading higher-order evidence, it can in fact be rational to hold an akratic combination of beliefs. For those who want to resist that conclusion, the challenge is to explain why akratic combinations of beliefs are not the correct way to reflect one's evidence, but this has proven difficult. In this final section, I argue that resilience offers an attractive explanation.

I will focus on a case of apparently rational epistemic akrasia developed by Allen Coates (2012), in which it is especially clear how resilience helps resolve the problem:

“Watson is an apprentice of Holmes, a master sleuth. As part of his training, Watson will often accompany Holmes to crime scenes and other locations, size up the evidence as best he can, and tell Holmes what conclusion he has drawn, and how he has drawn it. Holmes will then assess Watson’s conclusion as rational or irrational, though not as true or false. Of course, this assessment is based in part on whether Holmes thinks the evidence supports the conclusion. But just as a logic student may use invalid steps to arrive at a conclusion that follows validly from the premises, so too Watson may use poor reasoning to arrive at a conclusion that is nevertheless supported by the evidence. In such a case, Watson would be irrational in holding his conclusion, and Holmes will assess it accordingly. Thus, it is possible for Holmes to arrive at the same conclusion from the same evidence as Watson, and still claim that Watson’s belief is irrational. Watson is aware of this, and so he cannot infer from such a claim that Holmes thinks his conclusion is false. In fact, he cannot even infer that Holmes thinks that the evidence does not support his conclusion. All he can infer is that Holmes thinks that he has arrived at his conclusion irrationally. This is by design: Watson is to consider the evidence on his own until he arrives at a conclusion rationally, and his knowing Holmes’ own conclusion would interfere with this. Now suppose that Holmes brings Watson to a crime scene, that the evidence indicates that the butler is guilty, and that Watson uses good reasoning to arrive at that conclusion. In short, Watson rationally believes that the butler did it. But when he tells Holmes of his conclusion and how he arrived at it, Holmes only response it, ‘Your conclusion is irrational.’ Since Holmes is a master sleuth, Watson is justified in believing Holmes to be correct: Holmes’s testimony on these matters is very authoritative. But authoritative though he is, he is not infallible, and this is one of the rare occasions in which he is wrong. So when Watson accepts Holmes’s assessment, he accepts a falsehood. Watson, then, may reasonably but wrongly judge that his conclusion is irrational. Therefore, if he nevertheless maintains his belief in the butler’s guilt, both it and his epistemic judgment of it are rational. Yet, in holding them both, he is akratic.” (Coates 2012: 114)

Since both of Watson’s beliefs are rational in light of his evidence, this seems like a *prima facie* case of rational epistemic akrasia. And several authors do in fact accept this possibility, Coates included.<sup>12</sup> This, however, is highly counterintuitive, and leads to other problems concerning how one should reason from the akratic beliefs, what actions they rationalise, and more.<sup>13</sup> But the alternatives to allowing rational akrasia are problematic as well. If Watson adopts the ‘steadfast’ approach, as some philosophers recommend, and believes that the butler did it and that his evidence supports this, Watson fails to acknowledge his higher-order evidence. But if Watson instead adopts the ‘conciliatory’ approach recommended by others, and becomes less confident that the butler did it and that his evidence supports this, he will fail to acknowledge his first-order evidence. So either way, Watson will fail to respect some of his evidence.<sup>14</sup>

How does resilience help us understand cases such as this? A crucial feature of the above case is that Holmes’s testimony isn’t evidence of the butler’s innocence. Holmes may judge Watson’s belief to be irrational, even if he thinks that it is true. As Coates observes, without this feature it would clearly not be rational of Watson to continue to believe in the butler’s guilt. The significance of this becomes clear when we consider Watson’s beliefs in terms of the underlying credences. Suppose that the evidence at the crime scene supports believing that the butler did it partly in virtue of supporting a credence in this above the threshold for categorical belief. After correctly assessing the evidence, Watson thus adopts the belief that the butler did it, partly in virtue of adopting a high credence in the butler’s guilt. When Watson hears Holmes’s testimony, this makes it rational for Watson to believe that his belief in the butler’s guilt is irrational. Again, it is natural to interpret this in terms of

<sup>12</sup> For other defenders, see e.g. Wedgwood (2012), Lasonen-Arnio (2014), and Williamson (2014).

<sup>13</sup> For criticism, see e.g. Horowitz (2014).

<sup>14</sup> For the ‘steadfast’ approach to misleading higher-order evidence, see e.g. Titelbaum (2015). For the ‘conciliatory’ approach, see e.g. Feldman (2005), Christensen (2010b), and Horowitz (2014).

credence, namely as the belief that his high credence in the butler's guilt is irrational. But since Holmes's testimony isn't evidence of the butler's innocence, it doesn't by itself support moving to a lower credence in the butler's guilt. Recall that as the case is set up, Watson's credence may be correct in light of the evidence, even if it is irrational. This makes it plausible that Watson may maintain his high credence, even after hearing Holmes's testimony. However, Watson's belief that his high credence is irrational should have the effect of lowering the resilience of that credence. It should dispose him to revise his credence more readily in light of new first-order evidence about the culprit, and in light of new higher-order evidence that, contrary to Holmes's testimony, does suggest that a different level of credence is correct. But in that case, Watson's credence will no longer qualify as a categorical belief, since this requires a certain level of resilience (assuming the stability theory of belief).

In this way, Watson can respect both his first- and higher-order evidence, without it being akratic. His first-order evidence is correctly reflected in his high credence in the butler's guilt. His higher-order evidence is correctly reflected in his low confidence in his own credence being rational in light of the evidence, and in the correspondingly low resilience of that credence. And it is not a case of akrasia because, while believing  $p$  is rationally incompatible with believing that believing  $p$  is irrational, it is not necessarily irrational to hold some level of credence while being doubtful that that level is rational. If Watson responds to his evidence as he ought to do, he is therefore not epistemically akratic.

As in the discussion of conditionalization, this solution faces the potential objection that, while it may work in cases where the higher-order evidence doesn't suggest that a different credence level is correct, it doesn't work when this condition isn't met, i.e. in lopsided cases. But it is far from clear that such cases are problematic to begin with. Coates clearly thinks that his verdict that Watson's akratic beliefs are rational depends on Watson not being able to infer from Holmes's testimony that the butler is innocent. Had he been able to infer this, his belief that the butler is guilty would not be rational, and hence there wouldn't be a puzzling case of rational akrasia. At least some other cases of supposedly rational epistemic akrasia in the literature appear to be non-lopsided in this way as well.<sup>15</sup> But I shall have to leave detailed discussion of this objection for another occasion.

## 6. Conclusion

I have argued that higher-order defeat should sometimes be understood as undermining the resilience of one's credences rather than their level. I showed how this integrates with a general picture of higher-order evidence as something that constrains one's higher-order credences, and outlined how the proposal helps explain two puzzling features of higher-order defeat. Obviously much more can and must be said to develop and motivate this account. But I hope to have showed that the account is sufficiently promising to warrant further work.

**Acknowledgements:** This paper has been presented in Edinburgh, Aarhus, and Roskilde. I am grateful to those present for fruitful discussion. Thanks in particular to David Christensen, Mattias Skipper, and Jens Christian Bjerring, who read earlier drafts and provided helpful comments.

---

<sup>15</sup> For example, Wedgwood (2012) discusses higher-order evidence pointing to one's own irrationality or incompetence, but not evidence supporting that a different inference or conclusion is in fact correct.

## References

- Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2), 187-217.
- (2010a). Rational Reflection. *Philosophical Perspectives*, 24(1), 121-140.
- (2010b). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81(1), 185-215.
- Coates, A. (2012). Rational Epistemic Akrasia. *American Philosophical Quarterly* 49 (2): 113-24.
- Egan, A., & Elga, A. (2005). I Can't Believe I'm Stupid. *Philosophical Perspectives*, 19(1), 77-93.
- Feldman, R. (2005). Respecting the Evidence. *Philosophical Perspectives* 19 (1): 95–119.
- Foley, R. (1994). *Working Without a Net*. Oxford University Press.
- Frankish, K. (2009). Partial Belief and Flat-Out Belief. In F. Huber & C. Schmidt-Petri (eds.), *Degrees of Belief*, Springer, 75-93.
- Hansson, S. O. (2008). Do We Need Second-Order Probabilities? *Dialectica*, 62(4), 525-533.
- Horowitz, S. (2014). Epistemic Akrasia. *Noûs*, 48(4), 718-744.
- Joyce, J. M. (2005). How Probabilities Reflect Evidence. *Philosophical Perspectives*, 19(1), 153-178.
- Keynes, J.M. (1921). *A Treatise on Probability*. London: Macmillan.
- Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88 (2): 314-345.
- Leitgeb, H. (2014). The Stability Theory of Belief. *Philosophical Review*, 123(2), 131-171
- (2015). I—The Humean Thesis on Belief. In *Aristotelian Society Supplementary Volume* (Vol. 89, No. 1, pp. 143-185).
- (2017). *The Stability of Belief. How Rational Belief Coheres with Probability*. Oxford University Press.
- Loeb, L. E. (2002). *Stability and Justification in Hume's Treatise*. Oxford University Press.
- Popper, K. (1972). *The Logic of Scientific Discovery*. Hutchinson.
- Rasmussen, M. S., Steglich-Petersen, A., & Bjerring, J. C. (2018). A Higher-Order Approach to Disagreement. *Episteme* 15: 80-100.
- Skyrms, B. (1977). Resiliency, Propensities, and Causal Necessity. *The Journal of Philosophy*, 74(11): 704-713.
- Steglich-Petersen, A. (2006). No Norm Needed: On the Aim of Belief. *The Philosophical Quarterly* 56(225): 499-516.
- (2009). Weighing the Aim of Belief. *Philosophical Studies* 145(3): 395-405.
- Titelbaum, M. (2015). Rationality's Fixed Point (Or: In Defense of Right Reason). *Oxford Studies in Epistemology* 4: 253-294.
- Wedgwood, R. (2012). Justified Inference. *Synthese* 189 (2): 1-23.
- White, R. (2009). On Treating Oneself and Others as Thermometers. *Episteme* 6: 233-50.
- Williamson, T. (2014). Very Improbable Knowing. *Erkenntnis* 79 (5): 971-999.
- Worsnip, A. (2014). Disagreement About Disagreement? What Disagreement About Disagreement? *Philosophers' Imprint*. 14: 1-20.