Instrumentalism and Higher Order Defeat[1]

Abstract: Standard accounts of higher order defeat run into well-known Bayesian hurdles. In this paper, I explore the prospects of overcoming those hurdles by applying a thoroughgoing instrumentalist approach: treating our reasoning capacities as we'd treat any other instrument. I argue that instrumentalism cannot coherently accommodate the defeatist's position when that position involves genuine higher order uncertainty—that is, rational uncertainty about what rationality requires. At the end of the paper I gesture at an approach to modeling our responses to defeaters that is both instrumentalist and Bayesian, but that abandons the thought that the cases typically described as cases of higher order defeat involve genuinely higher order uncertainty.

1. Introduction

The "you just believe that because" challenge[2] sounds something like this: "You're just an atheist because you were raised by atheists," "you're just an externalist because you did your PhD in Oxford," "you just think that death is bad because you've been programmed to think so by natural selection." Learning such facts about our beliefs' causal history can be disturbing, and it sometimes seems appropriate to abandon the beliefs that have been influenced in problematic ways. This form of challenge to our beliefs is an instance of what is sometimes called "higher order defeat." In the cases of higher order defeat that I'll focus on in this paper, we gain evidence that bears (in an unflattering way) on the causal process by which we evaluated our evidence.

Several authors have pointed out that the view which recommends abandoning belief in response to higher order defeaters (let's call it "defeatism") isn't easily captured by standard Bayesian frameworks (Christensen (2010), Schoenfield (2018), Horowitz (2019), Bradley (2021)

[1] For helpful discussion and feedback, special thanks to Jennifer Carr, Sinan Dogramaci, Cian Dorr, Kevin Dorst, Hartry Field, Sophie Horowitz and Susanna Rinard. Thanks also to audiences at the London School of Economics, New York University, Syracuse University, University of Chicago, University of Helsinki, University of Maryland, and University of Warwick.

[2] I borrow the name of this widely discussed challenge from White (2010). For a smattering of approaches to the topic see also Cohen (2000), Street (2006), Christensen (2010), Avnur and Scott-Kakrues (2015), Vavova (2018), Pittard (2019), Srinivasan (2019), Schoenfield (2022), and Elga (ms.)

and Levinstein (ms.)[3]). On a very abstract level, the problem is that, according to the typical treatment of higher order defeat, there is a conflict between the defeatist's recommendation about how to respond to evidence that we've reasoned irrationally (abandon your belief!) and the recommendations of the agent's prior credences (stick to your guns!). This conflict in recommendations exists because, on the standard picture, the agent's prior (or, if you prefer, the rational ur-prior[4]) is both opinionated about what the evidence supports, and regards facts about the agent's later cognitive misadventures, as largely irrelevant to the question of what the evidence supports.

Why does higher order defeat face this challenge, when the Bayesian has no problem capturing other forms of defeat? The reason, I'll argue, is that the Bayesian story of higher order defeat is internally conflicted about the role that reasoning plays in justifying our views about the evidential support relations. One way out of the conundrum is to adopt what is sometimes called an "instrumentalist" approach to our reasoning processes. In what follows I'll explain why the instrumentalist approach, which works so well in other cases of defeat, can't apply to higher order defeat so long as we think of higher order defeat as being genuinely higher order – that is, as involving rational uncertainty about what's rational. At the end, I'll gesture towards an instrumentalist model of what agents are doing when they respond to evidence about their reasoning processes, which avoids these problems by abandoning the 'higher order' aspect of higher order defeat.

2. The Hypoxia Case

Let's start with a toy case. The case isn't intended to be particularly realistic, but I use it because it's easy to apply formalisms to:

> HYPOXIA (a variant of a case from Elga (ms.)): You're a cargo pilot who was hired to fly from London to Philadelphia to deliver some goods. While in the air, you get a message from your employer. They inform you that, due to some recent developments, it would be optimal for the goods to be delivered to Los Angeles rather than Philadelphia.

[3] Weisberg (2015) is also relevant here, though his focus is not on higher order defeaters.

[4] The rational ur-prior has been understood in different ways by different authors, but, approximately, it is the credence function that is rational in the absence of evidence. See Meacham (2016) for more detail.

Unfortunately, they can't afford a refueling stop, but they're wondering whether you have enough fuel to safely make it to Los Angeles. At this point [t_0] your credence that you have enough fuel to make it to Los Angeles (call this proposition **L**) is 0.5.

At t_1 you learn **E** (perhaps some information from the dials and gauges). You are rationally certain that E either supports a high degree of confidence in L or an equally high degree of confidence in ~L[5] You engage in some reasoning and rationally become confident that L on the basis of E.

At t_2, you get a call from ground control. They alert you that for the past several minutes (between t_1 and t_2) you've been suffering from hypoxia: a condition that impairs one's ability to evaluate evidence like E, even though everything seems normal from the inside. Pilots with hypoxia, you are told, do no better than chance at evaluating such evidence.[6] Call the information you get from ground control "**H**." (As a matter of fact, we can suppose, you don't have hypoxia, and have reasoned well, but you have every reason to trust the folks from ground control).

Defeatist verdict: Your credence in L after you learn H, should go back to 0.5 (or thereabouts).[7]

In this case, you gain information that casts doubt on your ability to rationality evaluate evidence E. I'll call evidence that casts doubt on your ability to rationally evaluate evidence "**a higher order defeater**."[8] Let's call the proposition that believing L is rational in response to

---

[5] More accurately, E combined with your background knowledge supports a high credence in L or a high credence in ~L. I'll leave the background knowledge implicit from hereon.

[6] As a heuristic, you can imagine that they have a fair coin in their head that gets flipped that determines whether they end up taking the evidence to support L or to support ~L.

[7] For defenses of such verdicts see e.g. Christensen (2010, 2019), Horowitz and Sliwa (2015), Vavova (2018), Pittard (2019), Bradley (2021) and Elga (ms.). For arguments against defeatist verdicts see e.g. White (2010), Kelly (2011), Lasonen Aarnio (2014), Titelbaum (2015), Srinivasan (2019) and Weatherson (ms.).

[8] I'll be assuming for now that the impairment in question impacts your ability to determine which conclusion is *rational* given your first order evidence. This is a common assumption in the debates about higher order evidence. (Christensen (2010), for example, says that higher order evidence "rationalizes a change of belief precisely because it indicates that my former beliefs were rationally sub-par" (p.185). In a similar vein, Dorst (2019) claims that taking peer disagreement seriously only makes sense if learning about the disagreement gives us evidence that we've reasoned irrationally). However, Christensen (2016) and Kopec and Titelbaum (2019) defend the idea that what

evidence E "**RatEL**." The view I'll be calling **higher order defeatism** is one according to which, in HYPOXIA, you ought to reduce your confidence in RatEL when you gain evidence suggesting that your ability to evaluate the evidence is impaired, and your rational reduction of confidence in RatEL is what explains your rational reduction of confidence in L.

3. Strange Things About Higher Order Defeat

As I mentioned in the introduction, in a classical Bayesian setting, the story the defeatist wants to tell in HYPOXIA runs into troubles, which I think are best highlighted through the lens of accuracy considerations. These have been discussed at length elsewhere, so for now I'll just enumerate some of them.

(1) It's often assumed that the rational way to revise our probabilities in response to evidence is to adopt the probabilities that maximize expected accuracy from the perspective of the prior probability function (Greaves and Wallace (2006)). However, if p is the rational probability function prior to boarding the plane (or the rational ur-prior), the procedure which maximizes expected accuracy relative to p, is one that recommends that if you find yourself in the situation described in HYPOXIA, you *ignore* the higher order evidence, and maintain your high confidence in L. (Christensen (2010), Schoenfield (2018), Levinstein (ms.))[9]

really matters in these cases is the question of whether you *reliably*, rather than *rationally*, evaluated the evidence. For my purposes, it doesn't matter whether we think of the relevant impairments as bearing first and foremost on the rationality or the reliability of your method of evaluating evidence. Because much of the higher order evidence literature focuses on evidence regarding a lack of rationality, principles used in this literature (like the "bridge" principles I'll discuss later) are stated using rationality instead of reliability. If one goes the reliability route, one will need similar bridge principles regarding reliability rather than rationality, but as long as those are substituted in consistently, the arguments in this paper will remain unchanged.

[9] Conditionalization is famously motivated by the idea that it is the plan that maximizes expected accuracy (Greaves and Wallace (2006)). What the defeatist is recommending does indeed look like a deviation from conditionalization. I focus on expected accuracy, however, because one might claim that the defeatist response is compatible with conditionalization if the proposition in question is essentially indexical (Christensen (2010)). Whether it can in fact be well represented this way gets into the weeds of debates about how to account for indexical information when updating. But regardless of whether the revision can be categorized as an instance conditionalization, it is a revision that is incompatible with some of the primary *motivations* for conditionalization. (While I appealed to the Greaves and Wallace argument for conditionalization based on expected accuracy, the same holds for the Briggs and Pettigrew (2020) accuracy dominance argument, as well as Dutch book arguments for conditionalization).

(2) In higher order defeat cases, you expect to be more accurate having the belief state supported by a subset of your evidence, than the belief state supported by your total evidence. (Schoenfield (2016)).

(3) Higher order evidence is predictably misleading – before learning whether or not you're impaired, you should prefer (from an accuracy-perspective) the attitude that's rational given your current evidence, to the one that would be rationalized were you to find out whether you're impaired. (Horowitz (2019), Levinstein (ms.)).

Without going into too much detail, you can get a sense for why these troubles arise by thinking about the pilot's attitude before she boards the plane. A Bayesian will think that if E rationalizes a high degree of confidence in L, then the rational credences to have before boarding the plane, p, will be such that $p(L|E)$ = high. But notably, before boarding the plane, the pilot doesn't regard her cognitive condition while on the plane as having any bearing on how likely there is to be enough fuel to make it to Los Angeles given the readings of dials and gauges. In other words, conditional on E, she regards L as independent of propositions about the goings on of her brain while on the plane. Then, once she's actually on the plane, she *does* take propositions about her impairment to be relevant to L. This leads to all manner of trouble, as one would expect when one fails to conditionalize in the usual ways.

4. Instrumentalism

Learning that one's vision is unreliable, that one's friend is unreliable, or that one's thermometer is unreliable, can, in principle, be captured perfectly well in a Bayesian setting – with no oddities like the ones described above. What goes wrong when we cast doubt on our ability to *reason* well? In the following sections I'll explain why the mechanism by which classical Bayesianism can easily capture undermining defeat in general won't work for higher order defeat.

So let's first look at how a Bayesian can accommodate other forms of undermining defeat. Consider this example:

VISUAL IMPAIRMENT: You currently assign a 0.5 credence to the proposition that the wall in the room next door is red (R). You enter the room and the wall appears red to you (AR).

> As a result, your credence that the wall is red goes up significantly. You are then told that something is wrong with your color vision, and that the objective chance of you judging correctly whether a wall is red under these circumstances is 0.5 (call this proposition D).[10] Your credence that the wall is red goes back down to 0.5 (or thereabouts).

A simple Bayesian story of this case could go as follows: Before seeing the wall, your probability function *p* is such that:

p(R)= 0.5
p(R|AR) = high
p(R|AR&D) = 0.5 (or thereabouts).

Straightforward conditionalization will yield the verdict that after seeing the wall and hearing the testimony about your visual impairment, your credence should be 0.5.

The way I modeled the situation above is "**instrumentalist**" – it involves treating the outputs of our mental processes as "readouts" of an instrument. When we treat an output of a mental process as an instrument, we conditionalize on a proposition about the readout of the instrument, and end up with a posterior credence that accounts for both our prior probability in the proposition in question, and our views about the instrument's reliability.

Let's apply this thought to a specific case where you have an instrument that produces one of two readings: {P,~P}. Assume that you regard the instrument's reliability as the same regardless of which of P or ~P is true. (In other words, the instrument isn't more likely to produce an accurate reading conditional on P than it is conditional on ~P or vice versa). Let *r* be the credence you assign to the instrument reading P conditional on P (which equals the credence you assign to the instrument reading ~P conditional on ~P). Finally, let p_D be a credence function that has learned of a defeater and so thinks of the instrument in question as performing no better than chance. When the instrument performs no better than chance at determining whether P, the value of *r* is 0.5, and it follows from Bayes' Theorem that:

---

[10] If you're skeptical of the use of objective chance here, feel free to substitute D with any proposition P concerning your visual capacities which is such that your prior probability function, Pr, is such that Pr(Red|It appears to you that the wall is red & P) = 0.5. (Perhaps, for example, P contains information about track records of people forming redness judgments in circumstances like yours).

$$p_D(\text{P}|\text{Instrument reads “P”}) = p_D(\text{P})*r \,/\, [p_D(\text{P})*r + (1\text{-}p_D(\text{P}))*(1\text{-}r)] = p_D(\text{P})$$

In other words, if you know that your instrument, which will output either “P” or “~P”, is no more reliable than chance, then your credence in P upon learning the instrumental readout should be the same as your credence in P *prior* to learning the instrumental readout.

In the case of visual perception above, if, before seeing the wall, you start out 0.8 confident that the wall is red, you have a visual experience as of a red wall, and are then told that your vision does no better than chance at detecting which of {R,~R} is true, your credence that the wall is red should end up back at where you started: 0.8. If you started out at 0.4, it should end up at 0.4, etc. On the instrumentalist Bayesian story, the effect of the undermining defeater is to “undo” the credal revision that happened through a process you now regard as defective. (See also White (2009), Pittard (2019) and Isaacs (2021)).

It’s worth noting that some philosophers are unhappy with an instrumentalizing approach when it comes to perception (e.g. Williamson (2000) Pryor (2000)). If the instrumentalizing approach is problematic in the case of perception, then it will turn out that perceptual defeat *also* poses worries for classical Bayesianism (Pryor (2013), Weisberg (2015), though see also Miller (2016, 2017)). But regardless of your views about the instrumentalist approach to perception, the general point is this: Bayesian models of (undermining) defeat work well when we can represent our situation instrumentally. Many instances of undermining defeat *are* well-modeled instrumentally (if you don’t like the application to vision, try testimony). I want to explain what prevents us from treating our judgments about what the evidence supports instrumentally.

5. Applying Instrumentalism to HYPOXIA

Let’s try to tell an instrumentalist story about HYPOXIA. Recall, at this point I am assuming that the immediate impact of higher order defeaters is to make you uncertain about a higher order proposition – a proposition about what your evidence supports. If the instrumentalist story is meant to apply to uncertainty about the higher order proposition, then you must be gaining evidence that your “instrument” does not reliably determine what the evidential support relations are. If we were to straightforwardly apply the instrumentalist story of defeat to a case where you learn that you are unreliable about higher order propositions, like HYPOXIA, the story would go as follows:

Step 1: You start with a certain credence in RatEL(the proposition that E supports L).[11]

- p(RatEL) = a

Step 2: You receive E, do some reasoning, and obtain an "instrumental readout" concerning whether E supports L. We might think of the readout as a judgment[12] or "intellectual seeming" concerning what E supports. I'll use "J" (for judgment) as a placeholder for whatever proposition most appropriately expresses the fact that your cognitive instrument's output was "RatEL."

Step 3: You conditionalize on J and boost your credence in RatEL.

- p(RatEL|J) = b>a

Note that in the hypoxia story you've also learned E by the time you've judged what E supports, but because E isn't itself evidence for or against RatEL, it's not playing a role in justifying the credal boost to RatEL. Nonetheless, it's also true that p(RatEL|E&J) = b>a.

Then, if a defeater like H shows up, which tells you that your judgments perform no better than chance, the evidential impact of your judgment will be undermined, and so:

Step 4: You conditionalize on H. Your credence in RatEL will return to where it was in Step 1.

- p(RatEL|J&H) = a

(And for the same reasons as above, it's also true that, p(RatEL|E&J&H) = a)

So far, so good, and so Bayesian. But what's going on with your credences in L (the first order proposition) throughout all of this? The relationship between first order credences and higher order credences that higher order defeatists tend to endorse is described by what are called

[11] It will be simplest in what follows to think of your starting credence function as the rational ur-prior, though nothing essential rests on this.

[12] Horowitz and Sliwa (2015) think of an agent's judgment as the proposition they are disposed to be most confident in on the basis of the first order evidence alone. See their paper for a more detailed discussion of the role that judgments play in higher order defeat.

"bridge principles." The bridge principles will ensure that, as your credences in RatEL fluctuate, your credences in L given E come along for the ride. The weakest and least controversial such bridge principle that gives the defeatist what they need is a version of Elga's (2012) "New Rational Reflection" or Dorst's (2019) "Hifi". I've modified the principle slightly so that it better accounts for the defeatist's verdict in HYPOXIA:[13]

> **BRIDGE:** If p is the rational ur-prior, $p^* = p(\cdot|p \text{ is the rational ur-prior})$, and $\text{Exp}_q(r)$ is the expectation of r, relative to a probability function q, then, for any body of evidence E,
>
> $$p(\cdot|E) = \text{Exp}_{p(\cdot|E)}(p^*(\cdot|E))$$

Given some supplementary background assumptions about the case (described in the appendix), we can derive the following:

> **Result**: If BRIDGE is true, then it follows from probabilism and conditionalization that, in HYPOXIA, you should assign a 0.5 credence to L once you learn H if and only if you should assign a 0.5 credence to RatEL once you learn H.

Result holds independently of instrumentalism, but it's important in this context because it is needed for applying instrumentalism to HYPOXIA. (While I've relegated the proof to an appendix, I recommend it, since it clarifies how the bridge principle works).

What if, instead of a proposition like L, you were reasoning about a proposition with a prior higher than 0.5? Suppose, for example, that you were initially (rationally) 0.8 confident in

---

[13] The difference between BRIDGE and NRR/Hifi is that, while BRIDGE tells you that your credences (when your evidence is E) should equal your expectation of the rational credence function that knows both E and which credence function is *the rational ur-prior*, NRR says that your credences should equal your expectation of the rational credence function that knows both E and which credence function is *rational given* E. The problem with appealing to NRR in a defeat context is that, in HYPOXIA you're meant to become uncertain, upon learning H, about what it is rational to believe *given E*. But while you may not be in a position to be confident about what E supports, you may well be in a position (because you know that defeatism is true) to be confident about what $E\&H$ supports (namely, a 0.5 credence in L). If that's right, then once you have H in hand, you may have no uncertainty at all about what's rational given your *total* evidence. Thus, if H is meant to change your expectation of the value of some credence function, and this change in expectation is meant to change your credence in L, it had better be a credence function that knows what's rational given E, and not (only) what's rational given E&H. Thanks to Kevin Dorst for discussion on this point.

some other proposition, call it M. You do some reasoning on the basis of your evidence E, judge that it supports M, and then discover that your reasoning processes are impaired (that is, they perform no better than chance at determining whether E supports M or E supports ~M). In this case, the defeatist will presumably recommend that your credence in M once you learn H, should go back to your prior probability in M of 0.8. (There's no reason it would go all the way down to 0.5). The defeatist verdict about the case, in combination with BRIDGE and some plausible background assumptions will yield the result that the agent's credence that E supports M, should be *greater* than 0.5, upon learning H.[14]

In other words, on the defeatist's picture, how confident you should be in the higher order proposition when you're told that you've been hypoxic will partly depend on what your prior in the *first order* proposition was. If your prior was 0.5 (as it was in our original case), your credence in the higher order proposition upon learning that you're hypoxic must also be 0.5 to get the defeatist's verdict. If, however, your prior in the first order proposition was *above* 0.5, your credence in the higher order proposition about the evidential support relations will have to also be above 0.5 as well. While this may seem initially surprising, it's exactly what we should have hoped for. If you have a high prior probability in M, then you should take your evidence, E, to also be evidence for the proposition *that E supports M*. After all, if M is probably true, you're more likely to get evidence that supports its truth. So if you get evidence E, that's some evidence that E is evidence for M.[15]

So here is where we are: Instrumentalism about higher order reasoning, combined with BRIDGE, seems to provide a satisfying account of how our first order and higher order credences evolve in response to a defeater in a way that, at least on its face, is consistent with Bayesian

---

[14]Proof is left as an exercise to the reader, but just for illustration suppose that the ur-prior, p assigns non-zero credence to two theories about p^*:

According to RatEM: $p^*(M|E) = 0.99$, and $p^*(M|{\sim}E) = .17$
According to ~RatEM, $p^*(M|E) = .17$ and $p^*(M|{\sim}E) = .99$.

We can now use the defeatist's verdict about this case (that you should return to your prior of 0.8 in M), BRIDGE, and the supplementary assumptions in the appendix to calculate how the defeatist should think about your credence in RatEM upon learning E, J and H.

(1) $p(M|EJH) = 0.8$ (Defeatist verdict)
(2) $Exp_{p(\cdot|EJH)}[p^*(M|EJH)] = 0.8$ ((1) and Bridge)
(3) $p(RatEM|EJH)(.99) + p(\sim RatEM|EJH)(.17) = 0.8$ ((2) and supplementary assumptions from Appendix)
(4) $p(RatEM|EJH) \sim .77$ (3)

[15] See Dorst and Hedden (2022) for a general explanation of why our first order evidence is often also evidence about what the evidence supports.

conditionalization, and the Bayesian's usual treatment of defeat. This is a good juncture to zoom out for a moment and think about the crucial insight the instrumentalist offers and why defeatists have generally wrinkled their nose at it.

The instrumentalist model I described above treats the rational ur-prior in much the way we'd treat any other expert probability function (a more informed probability function, the objective chance function, etc.). You start out uncertain about what the expert thinks (in our case, you start out uncertain about RatEL). You then gain some evidence which justifies a revised opinion about the expert's probabilities (in our case, you learn J – a proposition about your 'instrumental readout'). You then align your probabilities in a prescribed way with the expert's (you follow BRIDGE). If you gain undermining evidence related to your views about the expert's probabilities (you learn H), you 'undo' the revision you made and return to your initial opinion.

One reason for the nose-wrinkling response to instrumentalism is that many epistemologists don't like the idea that our views about what the evidence supports are the result of conditionalizing on propositions about what I've been calling "judgments" – the outputs of our reasoning processes. While I think philosophers have been overly dismissive of instrumentalizing, there are some good reasons to be worried about it. For example, the idea that we go out into the world, gain evidence*,* and *then* conduct some empirical test ("form a judgment") to see what that evidence supports is arguably contrary to the spirit of the Bayesian idealization. The idealized Bayesian agent, as typically described, has done all the thinking, all the a priori work that needs doing, before setting so much as a toe on planet earth.[16]

This is a reasonable complaint against applying instrumentalism in the idealized Bayesian context, but notice also, that, for the same reason, it's at least a *bit* awkward to think of the idealized Bayesian agent as doubting their views about what the evidence supports as a result of learning that those views were formed while at risk of hypoxia. If the agent's views about the evidential support relations were formed at the beginning of their epistemic life, and aren't based on cognitive processes they engaged in while on the plane, why should concerns about the reliability of those

[16] Some have also suggested that instrumentalist views can lead to a regress (Staffel (2025), p.94-95, see also Lasonen Aarno (2014)). For you might think that upon learning the readout of your instrument, you'll then need a new instrument to determine what credences are supported by the original instrument. I don't share this concern, at least as applied to the current context. I see no problem with the proposal that we model certain forms of complex reasoning - the kind subject to higher order defeat - as instrumental readouts, without adopting the view that *any* inference involves conditioning on such a readout. Not all inferences are susceptible to higher order defeat. There are other issues that have been raised as well (see e.g. Enoch (2010), Horowitz and Sliwa (2015) and Schoenfield (2015)), but addressing the full cast of complaints against instrumentalism is beyond the scope of this work.

processes undermine their views?[17] In other words, dismissing instrumentalism on the grounds that reasoning done in response to evidence is irrelevant to the idealized Bayesian agent's views about the evidential support relations, sits awkwardly with the defeatist position. This tension is at the heart of the Bayesian troubles for the *non*-instrumentalist defeatist story. Given that, on the non-instrumentalist story, the initial opinions about the evidential support relation aren't based on the readouts of instruments, it's no surprise that the initial opinions recommend sticking to your guns *regardless* of any instrument malfunction. (After all, from the initial standpoint, the unreliable instrument is both unnecessary and irrelevant).

Note that, in contrast, on the instrumentalist's story, the agent must start out agnostic about whether E supports L. For recall that the defeatist will need the agent's credence in RatEL to be 0.5 after learning that they're hypoxic in order to get the desired verdict of a 0.5 credence in L. As shown earlier, if you boost your credence in response to the readout of an instrument, and then learn that the instrument is no more reliable than chance, conditionalization will have you return to your prior. So if you're meant to assign a 0.5 credence to RatEL after learning H, the instrumentalist must say that your prior in RatEL was 0.5. If you started out *more* (or less) than 0.5 confident in RatEL prior to boarding the plane, then learning about problems with the instrument shouldn't take you to 0.5, but to whatever your prior happened to be.[18] This points to another awkwardness with the non-instrumentalist account: The non-instrumentalist story usually assumes that you start out with a high credence in RatEL. This then raises the question: if you think a reduction of confidence in response to H is warranted, why a reduction to 0.5 specifically, if 0.5 wasn't your prior? The defeatist can't appeal to the fact that this is what's required to get the defeatist verdict about L, because the reduction of confidence in RatEL is meant to *explain* the reduction of confidence in L.

---

[17] This consideration might raise the following question: Why would we ever expect a theory of ideal rationality to work for agents who are meant to take their *irrationality* into account? But as Christensen (2007) has emphasized, we are not looking for a theory of ideal rationality to help out agents who are irrational. We're looking for a theory of ideal rationality that accounts for the fact that, *even ideal agents can get (misleading) evidence that they are irrational*.

[18] Here's another way to make the same point: We showed that to get the defeatist verdict, $p(\text{RatEL}|\text{EJH}) = 0.5$. If the only impact of H is to undermine the relevance of J (in other words, H has no evidential relevance to RatEL beyond its undermining effect), then $p(\text{RatEL}|\text{EJH}) = p(\text{RatEL}|\text{E}) = 0.5$. Since, in this case, E is not evidence for the proposition that E supports L, it follows that $p(\text{RatEL}) = 0.5$. As discussed previously, E *can* be evidence for whether E supports a proposition that you have an opinionated prior about. But given the setup of the case in which you are 0.5 in L and ~L, E will not be evidence for the proposition that E supports L.

The instrumentalist story, as we'll see, has its own problems, but before getting to those, let's first summarize the account from start to finish: You start out with a 0.5 credence in RatEL. You then form a judgment about what E supports and learn J: the proposition that you judged RatEL. Having conditioned on J, your credence in RatEL goes up (it doesn't matter exactly to where). You also learn E, and adopt a high credence in L in response to E (all perfectly aligned with your higher order views, as described by BRIDGE). Then you learn that the process producing your judgments about what the evidence supports is unreliable, and so your credence in RatEL goes back to 0.5, where it started. Your credence in L comes along for the ride, and there is no violation of conditionalization, no expecting more accuracy from less evidence, and none of the other oddities that the typical Bayesian story of higher order defeat usually encounters.

So why not adopt this lovely picture, which not only offers a way out of the Bayesian troubles, but also provides a unified picture, treating undermining defeat about higher order matters just as we treat undermining defeat about many other matters? In the next section I'll argue that there's a formal problem with instrumentalism about higher order defeat: the instrumentalist about higher order defeat has, by now, asked the notion of rationality to do more things than it consistently can.

6. Why We Can't Be Higher Order Instrumentalists

Here's the trouble: We've established that, on the instrumentalist picture, your prior in RatEL must be 0.5. But the bridge principle that the defeatist endorses will yield the result that if your prior in RatEL should be 0.5, your prior credence in L given E should also be 0.5. This is already problematic, since it was stipulated in the case that it was rational to be confident that L in response to E. But if your credence in L given E should be 0.5, and it's rational to conditionalize, how can we say that it's rational to be confident that L in response to E?

Perhaps the response to this is to simply abandon the initial stipulation that E supports L. If we are wholehearted instrumentalists, we might insist that E, all on its own, *doesn't* support L. We also need the instrumental readout, J. So on the new version of the view, we'll have:

$p(\mathrm{L}) = 0.5$
$p(\mathrm{L}|\mathrm{E}) = 0.5$
$p(\mathrm{L}|\mathrm{E\&J}) = \text{high}$

True - we've denied one of the stipulations of the case (that E supports L). But we still have something close to that initial stipulation, namely that *E&J* supports L. As long as we allow for the agent to be introspective, we can still tell roughly the same story. Our agent starts out 0.5 in L before boarding the plane. She gets on the plane, learns E, does some reasoning, forms some judgments on the basis of that reasoning (is aware of all this), and proceeds to boost her confidence in L. Then she learns H and her credence returns to 0.5. It doesn't sound *that* bad. And in making this switch we've reaped some significant benefits. This account is fully compatible with Bayesianism! No oddities.

But we now run into another difficulty: making sense of the content of the proposition being judged (J, recall is a proposition about the agent's judgment). We can no longer say that the agent judges that "RatEL." For if she has thought through the considerations in this paper, then she will be quite confident that E *doesn't* support L. And if you're confident that E doesn't support L, how can you judge that it does? Additionally, why would you have a 0.5 credence in RatEL? On the instrumentalist picture under discussion, you should think that E supports neither L nor ~L, and so your credence in RatEL should be close to zero. But if you don't start out with a 0.5 credence in RatEL, and don't judge at any point that E supports L, then the entire instrumentalist picture of how we come to adopt opinions about what the evidential support relations are falls apart.

The reason, then, that we can't model higher order defeat in the way we model other forms of defeat is this: On the one hand, we need the *content* of our instrumental readouts to be a meaty claim about what it's rational to believe in response to E, thereby presupposing that E, all by itself, rationalizes some view concerning L. On the other hand, instrumentalism is a way of thinking about the rational *use* of the instrument, and as a result (when combined with defeatism and the bridge principle) requires that E, on its own, supports neither L nor ~L. Top it all off with the agent's awareness of these facts and you're cooked.[19]

7. A Non-Higher-Order Approach to "Higher Order" Defeat

What should we take away from the fact that we can't instrumentalize our judgments about what the evidence supports? The traditional defeatist may take the arguments in this paper to be

[19] See Pittard (2019) for a fascinating discussion of the limits of instrumentalizing that differs from the one I'm describing here though is also, I believe, relevant to these cases.

more grist for the nose-wrinkling mill. Treating our reasoning capacities as instruments was already deemed distasteful (at best), and if it's not just distasteful, but incoherent, so much the worse for instrumentalism. Sure, the non-instrumentalist approach to defeat has some other problems, but what's a minor deviation from classical conditionalization compared to the self-undermining mess that the instrumentalist ends up in? The answer, then, might be to rest content with the idea that a few adjustments have to be made to the typical paradigms to account for higher order defeat. It's not as if this would be the first time we needed to tweak Bayesian doctrine to accommodate sensible epistemology: Remember logical uncertainty? Information loss?[20] Notice, however, that the idea that we should accommodate logical uncertainty and information loss is often motivated by the desire to have a Bayesian-*ish* theory that can apply to nonideal agents. Adding higher order defeat to the list of anomalies the Bayesian has to contend with *could* be thought of in a similar way, but then we'd be giving up on the view that even ideally rational agents are susceptible to this form of defeat, a view many defeatists have insisted on (e.g. Christensen (2007)).

An alternative to rejecting the Bayesian instrumentalist model is to keep it, but instead reject the idea that what we've been calling 'higher order defeat' involves genuinely higher order uncertainty. One way to deal with this is to distinguish different notions of rationality. The instrumentalist ran into problems because rationality had to be both the notion that the instrument is making assessments *about*, and a set of norms governing the instrument's *use*. If, instead of using instruments to assess which attitudes were rational we were instead using instruments to assess which attitudes were, say, *shmational*, there would be no problem.

While we shouldn't proliferate notions of rationality without excellent reason, several philosophers have argued that accommodating higher order evidence requires multiple normative notions. Just to name a few examples: In the name of higher order defeat, Worsnip (2018)

[20] Levinstein (ms.) sees cases of higher order defeat as *instances* of information loss, like the case of Shangri-La (Artzenius (2003)). I'm not convinced that higher order defeat is best thought of as an instance of information loss. While a full argument is beyond the scope of this paper, the main reason I'm not convinced is that it's not easy to find a proposition that meets the following three conditions: (a) It was known before you learn H (b) It is no longer known after you learn H and (c) It's failure to be known after learning H explains a reduction of confidence in L. RatEL fails to meet condition (a). (The Bayesian defeatist doesn't typically think you start out certain that RatEL - even the rational ur-prior leaves space for the possibility that it is wrong about the evidential support relations). E fails to meet condition (b): nothing about the case suggests that you no longer know information about, say, where the dials and gauges are pointing once you learn H. And I'm not convinced that knowledge of what credences you had in the past can do the explanatory work, since knowledge of our previous credences is arguably not a precondition for rational updating - especially on a time-slice view of rationality.

distinguishes between substantive and structural rationality, Smithies (2022) distinguishes between ideal and nonideal rationality, Lasonen Aarnio (2021) distinguishes between dispositional and non-dispositional evaluations and I, in other work (2016), distinguish between the update procedure that's best to follow from the one that's best to plan to follow.

Staffel (2025) takes a different approach, distinguishing instead between two types of *attitudes*: terminal and transitional attitudes, and argues that this distinction can play a role in resolving higher order evidence's notorious difficulties. While it may seem like the last thing we need is another pair of notions attempting to fix higher order evidence's problems, I can't resist sketching a proposal that I find to be a particularly compelling instrumentalist model of cases like HYPOXIA. While some of the distinctions above *may* help with building a Bayesian account that applies to ideal agents, that's not what the model I'll sketch below does. At this point, I'm not trying to vindicate an idealized theory of Bayesian defeatism – I'm just seeking therapy. I want a model that explains what *I* am doing when I abandon belief in response to a higher order defeater. Like Staffel (2025), the proposal I'll sketch below (inspired by the literature on 'fragmentation'[21]), distinguishes between types of attitudes, instead of normative notions.

Start with the following observation: if a theorist of my psychology were asked, prior to my boarding the plane, "what is Miriam's credence in RatEL?" or "what is Miriam's credence in L conditional on E?" there are different ways they could respond. The theorist could say: "Credences are grounded in dispositions. Although Miriam has never consciously considered the question of whether E supports L, she has dispositions which make it appropriate to attribute to her a high credence that E supports L, or a high conditional credence in L given E." On the other hand, the theorist might say, "Let's ask her." If the theorist asked, and, within 5 seconds, I didn't have an answer, the theorist might say "It appears that Miriam doesn't know. She's agnostic about whether E supports L." There's nothing wrong with either answer, because the notion of credence is vague. We might sometimes be interested in what somebody is disposed to think and do upon careful and unlimited reflection, and sometimes be interested in how somebody is disposed to respond instantaneously. To fix terminology, let's say that, immediately prior to boarding the plane, I have a high *reflective-credence* in L given E, but I assign a 0.5 *instantaneous credence* to

[21] E.g. Elga and Rayo (2021)

L given E.[22] If we ask: what is Miriam's credence in L given E before boarding the plane *simpliciter*, the appropriate answer is: it's indeterminate.

Now I get on the plane and receive evidence E. Before doing any reasoning I might say things like this: "I have no idea whether E supports L, but it probably supports L if I conclude that it supports L, and it probably supports ~L if I conclude that it supports ~L." Then I engage in some reasoning, and become confident that E supports L. This process sure has an instrumentalist feel to it. But if I am instrumentalizing, why don't I run into the self-undermining mess that I described in the previous section? Answer: because my uncertainty on this model is not higher order.

Here is how to spell out the picture: My instantaneous-credence function (*i*) starts out agnostic about my reflective-credence function (*r*)'s views on the probabilistic relations between E and L. After I reason, my instantaneous-credence function learns, and conditionalizes on, a proposition about my reflective credences. Since, under normal conditions, *i* regards *r* as an expert, learning *r*'s opinion about L, changes *i*'s views about L. However, once *i* learns about the hypoxic conditions, *i'*s view about *r*'s expertise changes: *i* no longer thinks that my reflective credences on the plane are reliable and so returns to its original agnostic view about L.

What happens to my reflective credence function through all this? Before boarding the plane and before doing any reasoning, *r* already was opinionated about L given E. That is, *r* had the following conditional credence:

$r(L|E) = \text{high}$

Before the boarding the plane, *r also* thinks L is likely given E, and that it seems to me that L is likely give E under hypoxic conditions.

$r(L|EJH) = \text{high}.$

But once I'm the one plane, reason, and learn that I'm hypoxic, it's not just my instantaneous opinions that become agnostic. My reflective ones do too. For as long as I'm on the plane and subject to the hypoxic conditions, no amount of reasoning will convince me that L is likely to be true. This means that *r* doesn't evolve by conditionalization. Despite *r* earlier assigning $r(L|EJH) = \text{high}$, *r* does not adopt a high credence in L when the total evidence is EJH. Why is this?

[22] There may be some similarities between my notion of an instantaneous credence and Staffel's (2025) notion of a transitional attitude, but unlike her notion, I don't take instantaneous credences to be in any way aimed at transitioning into a more settled state. They're just descriptions of a dispositional profile that's elicited over a short time span.

To see why *r* doesn't evolve by conditionalization, note that if we want credence functions that evolve by conditionalization, and we're characterizing those credences dispositionally, we must be very careful about what dispositions are involved. For consider the following bizarre notion of credence:

Paper-credence: A person's paper credences are the opinions that would be elicited from them after they read all the pieces of paper within arm's reach.

Note that paper-credences won't evolve by conditionalization. Here's an illustration: Suppose that at 9am I am sitting a table with a newspaper on it that has a weather report predicting rain. Even if I haven't read the newspaper, I'll have a high paper-credence in the proposition that it will rain because that is the opinion that would be elicited had I read all of the papers within arm's reach. At 9:30 somebody has moved the newspaper out of arm's reach, and so my paper-credence in the proposition that it will rain will have changed despite there being no change in my evidence. My paper credences will therefore have failed to evolve by conditionalization. But there is nothing mysterious about that. Credences will only evolve by conditionalization if the elicitation conditions for the notion of credence at play don't involve gaining information.

The reason reflective credences don't evolve by conditionalization is the same as the reason my paper credences don't. My reflective credences describe opinions that would be elicited by engaging in reasoning, and engaging in reasoning gives me information about the output of such reasoning processes. Before boarding the plane, my reflective credences are opinionated in virtue of information they *would have* if reasoning took place. But in the hypoxia story, that reasoning never does take place, just as in the newspaper story, I don't ever look at the newspaper. What happens once I'm on the plane is that information about what the output of reasoning processes taking place on the ground would be, is no longer available to me through reflection (just as in the newspaper story, at 9:30, the newspaper is no longer within arm's reach). Since that information is no longer available through reflection, my reflective credences change – not through conditionalization but through a change in the availability of information.

This is far from a fully fleshed out account. But one thing in the account's favor is that it offers something to both the instrumentalist and the non-instrumentalist. For there is, on this view, a sense in which I start out agnostic about the probabilistic relationship between E and L (my instantaneous credence doesn't know what my reflective credence thinks), and a sense in which I start out opinionated about it (my reflective credence already has an opinion). Although this is just a sketch, it's meant to serve as a demonstration that we might be able to provide a satisfying Bayesian instrumentalist model of what we're doing when we respond to higher order defeat. This will not be a higher order model on which higher order defeat involves rational uncertainty about rationality itself. Instead, there will be multiple credence functions that serve different roles in modeling the agent and that can be uncertain about one another.[23]

8. Conclusion

In my view, the reason that the typical treatment of higher order defeat is inconsistent with the standard Bayesian model isn't the result of a minor technical issue, but of a serious tension in the theory. According to this treatment, the rational agent starts with firmly held views about the evidential support relation – views that, crucially, are not based on the outputs of any instruments. Nonetheless, learning that a certain instrument for assessing the evidential support relation turns out to be unreliable, gets the agent extremely fussed.

If we think that we should abandon our opinions upon learning that our reasoning capacities are compromised, a much more natural, and I'd argue, honest characterization of the situation, is one on which we acknowledge that our views about the evidential support relations weren't the starting point after all, but were formed on the basis of an application of those capacities. Unfortunately, a wholehearted instrumentalist treatment of higher order defeat can't work if we think that the norms governing our use of the instrument, and the norms that the instrument is trying to discover, are one and the same.

---

[23] You might wonder whether the reflective-credence function is susceptible to defeat. I'd suggest that the answer is either no, or that the question involves a category mistake, though to argue for that fully would be beyond the scope of this paper. The crucial point is that what I'm calling "the reflective-prior" is just a description of highly stable dispositions about the way the agent regards probabilistic relations between propositions. It's not something that evolves over time in response to minor cognitive blips. The other point to note is that since hypoxia is a condition applied to an *agent*, and not a credence function, it doesn't make sense to ask what the reflective-credence function's view is about L conditional on learning that *it* is hypoxic. We *can* however ask questions like this: how is the agent stably disposed to think of the probability of L conditional on EJH, where J&H describe an episode of reasoning while hypoxic. And the answer to that will be that the agent's stable disposition is to regard L as probable given E and facts about a certain episode of reasoning on a plane.

This leaves us with a choice point: we could abandon an instrumentalist approach to higher order defeat, or we could keep the instrumentalism, but abandon the view that the cases under discussion involve genuinely higher order uncertainty. I provided a rough sketch of one way to develop the latter strategy (thinking of our credences in the "instantaneous" sense, as assessing questions about our credences in the "reflective" sense), but the general approach that involves keeping instrumentalism, and rejecting higher-order-ism, is one that could be developed in several different ways. Treating our reasoning capacities as instruments might not turn out as badly as we initially thought.

Appendix: Proof of Result

The claim I'll be proving in this appendix is that, assuming probabilism, conditionalization, and some background assumptions that are listed below, if BRIDGE is true, then your credence in L once you learn H in HYPOXIA should be 0.5 if and only if your credence in RatEL once you learn H in HYPOXIA should be 0.5. I'll assume that your total evidence at the end of the HYPOXIA story is the proposition E&J&H (the first order evidence, the proposition that you judged that the first order supports L, and the proposition that you were at risk of hypoxia at the time of the judgment). I'll abbreviate this as "EJH." I include J because I'm using the result in the context of an instrumentalist account. However, the result holds even if J is omitted and you think your total evidence at the end of the story is just E&H. (Simply omit "J" in the proof throughout). As a reminder, here is BRIDGE:

BRIDGE: Let p be the rational ur-prior and let $p^*=p(\cdot|p \text{ is rational})$. Then $p=\text{Exp}_p(p^*)$ and, for any proposition, E, $p(\cdot|E)=\text{Exp}_{p(\cdot|E)}(p^*(\cdot|E))$.

Result: It follows from Bayesianism (probabilism+conditionalization), BRIDGE, and the assumptions below, that your credence in L upon learning EJH should be 0.5 if and only if your credence in RatEL upon learning EJH should be 0.5.

Assumptions

Assumption 1: You are certain throughout that for real numbers r and r^* both greater than 0.5 (a) $p(\text{L}|\text{E}) = r$ if and only if $p^*(\text{L}|\text{E}) = r^*$ and (b) $p(\sim\text{L}|\text{E}) = r$ if and only $p^*(\sim\text{L}|\text{E}) = r^*$.

Recall that we stipulated in the description of HYPOXIA that you are certain that the evidence E supports one of L or ~L, and whichever it supports, it supports to the same degree (that is, you are certain throughout that either $p(L|E) = r$ or $p(\sim L|E) = r$). Assumption 1 adds that this holds relative to p^* as well, and that p and p^* *agree* about which of L or ~L is probabilified by E (meaning, conditioning on the evidential-support facts doesn't shift their valence in this case).

Assumption 2: $p^*(L|EJH)=p^*(L|E)$ and you are certain of this throughout.

In other words: (you are certain that) *given certainty about the rationality facts*, information about judgments and hypoxic conditions are irrelevant to L conditional on E.

Proof from Left to Right: Let $p_{EJH} = p(\cdot|EJH)$ and assume that it's rational to be 0.5 confident that RatEL once you've learned EJH. By conditionalization, this means that

(1) $p_{EJH}(\text{RatEL}) = 0.5$.

By Assumption 2:

(2) You're certain that $p^*(L|E)=p^*(L|EJH)$.

From BRIDGE and (2) it follows that

(3) $p_{EJH}(L) = \text{Exp}_{p(\cdot|EJH)}\,(p^*(L|EJH)) = \text{Exp}_{p(\cdot|EJH)}\,(p^*(L|E))$

Let's now calculate $\text{Exp}_{p(\cdot|EJH)}(p^*(L|E))$. By (1), Assumption 1, and conditionalization, it follows that for $r>0.5$:

(4) $p_{EJH}(p(L|E) = r) = 0.5$ and
(5) $p_{EJH}(p(\sim L|E) = r) = 0.5$.

Also by Assumption 1

(6) You're certain that $p(L|E)=r$ iff $p^*(L|E)=r^*$.
(7) You're certain that $p(\sim L|E)=r$ iff $p^*(\sim L|E)=r^*$.

Given probabilism (specifically, the requirement that you assign equal credences to propositions you regard as equivalent), it follows from (4), (5), (6) and (7) that

(8) p_{EJH} (p^*(L|E)=r^*) = 0.5 and
(9) p_{EJH} (p^*(~L|E)=r^*) = 0.5.

By probabilism, p^*(~L|E)=r^*, if and only if p^*(L|E)=1-r^*.

So it follows from (8) and (9) that

(10) $\text{Exp}_{p(\cdot|EJH)}\, p^*(L|E) = (.5)r^* + (0.5)(1-r^*) = 0.5.$

By (3) and (10), it follows that $p_{EJH}(L) = 0.5$

Proof from Right to Left: Let $p_{EJH} = p(\cdot\,|EJH)$ and assume that $p_{EJH}(L) = 0.5$. Suppose for reductio that $p_{EJH}(RatEL) \neq 0.5$ and without loss of generality that it is greater than 0.5. Thus, we're assuming that

(11) $p_{EJH}(RatEL) = p_{EJH}(p(L|E) = r) = x$, for some x and some r both greater than 0.5.

From (11) and Assumption 1, it follows that

(12) $p_{EJH}(p^*(L|E)=r^*) = x$

Since the propositions in: $\{p^*(L|E)=r^*,\ p^*(L|E)=1-r^*\}$ form a partition (Assumption 1 and probabilism), it follows from (12) that

(13) $p_{EJH}(p^*(L|E)=1-r^*) = 1-x$

From (12) and (13) it follows that

(14) $\text{Exp}_{p(\cdot|EJH)}(p^*(L|E)) = xr^* + (1-x)(1-r^*).$

Since $x>0.5$, and $r^*>0.5$, it follows from (14) that

(15) $\text{Exp}_{p(\cdot|EJH)}(p^*(L|E)) > 0.5.$

From BRIDGE:

(16) $p_{EJH}(L) = \text{Exp}_{p(\cdot|EJH)}(p^*(L|EJH))$.

By Assumption 2:

(17) You're certain that $p^*(L|EJH)=p^*(L|E)$

It follows from (16), (17) and probabilism that:

(18) $p_{EJH}(L) = \text{Exp}_{p(\cdot|EJH)}(p^*(L|E))$

By (15) and (18), it follows that $p_{EJH}(L)>0.5$ contrary to our assumption that $p_{EJH}(L)=0.5$

Bibliography

Arntzenius, F. (2003).“Some Problems for Conditionalization and Reflection.” *Journal of Philosophy* 100(7): 356-370.

Avnur, Y. and Scott-Kakures, D. (2015). "How Irrelevant Influences Bias Belief." *Philosophical Perspectives* 29 (1).

Bradley, D. (2021). "Bayesianism and Self-Doubt." *Synthese* 199(1-2): 2225-2243.

Briggs, R. and Pettigrew, R. (2020). “An Accuracy Dominance Argument for Conditionalization.” *Noûs* 54(1):162-181.

Christensen, D. (2007). “Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals” In T.S. Gendler and J. Hawthorne (eds.). *Oxford Studies in Epistemology Volume 2*. Oxford University Press.

Christensen, D. (2010). “Higher Order Evidence.” *Philosophy and Phenomenological Research* 81(1): 185-215.

Christensen, D. (2016). “Conciliation, Uniqueness and Rational Toxicity” *Noûs* 50(3): 584-603.

Christensen, D. (2019). "Formulating Independence" In M. Skipper & A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays.* Oxford University Press Oxford University Press

Cohen, G.A. (2000). *If you're an Egalitarian, How Come You're So Rich*? Harvard University Press.

Dorst, K. (2019). "Higher Order Uncertainty." In M. Skipper & A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays.* Oxford University Press.

Dorst, K. and Hedden, B. (2022). "(Almost) All Evidence is Higher Order Evidence." *Analysis* 82(3):417-425.

Elga, A. (2012). "The Puzzle of the Unmarked Clock and the New Rational Reflection Principle" *Philosophical Studies* 164 (1):127-139

Elga, A. (ms.). "Lucky to be Rational."

Elga, A. and Rayo, A. (2021). "Fragmentation and Logical Omniscience." *Noûs* 56 (3):716-741

Enoch, D. (2010). "Not just a Truthometer." *Mind* 119 (476):953-997.

Hedden, B. (2015). *Reasons without Persons: Rationality, Identity and Time*. Oxford University Press.

Horowitz, S. (2014). "Epistemic Akrasia." *Noûs* 48 (4):718-744.

Horowitz, S. (2019). "Predictably Misleading Evidence." In Matthias Skipper Rasmussen and Asbjørn Steglich-Petersen (eds.) *Higher-Order Evidence: New Essays*. Oxford University Press.

Horowitz, S. and Sliwa, P. (2015). "Respecting all the Evidence." *Philosophical Studies* 17(11): 2835-2858

Isaacs, Y. (2021). "The Fallacy of Calibrationism." *Philosophy and Phenomenological Research* 102 (2):247-260.

Kelly, T. (2011). "Peer Disagreement and Higher Order Evidence." In A.Goldman and D. Whitcomb (eds.) *Social Epistemology: Essential Readings*. Oxford University Press.

Kopec, M. and Titelbaum, M. (2019). "When Rational Reasoners Reason Differently" In M. Balcerak-Jackson and B. Balcerak-Jackson (eds.) *Reasoning: New Essays on Theoretical and Practical Thinking*. Oxford University Press.

Lasonen Aarnio, M. (2014). "Higher Order Evidence and Limits of Defeat." *Philosophy and Phenomenological Research* 88 (2):314-345.

Levinstein, B. (ms.) "Higher Order Evidence as Information Loss."

Meacham, C. (2016). "Ur-Priors, Conditionalization, and Ur-Prior Conditionalization." *Ergo* 3 (17): 444-492.

Miller, B. (2016). "How to Be a Bayesian Dogmatist." *Australasian Journal of Philosophy* 94(4):766-780.

Miller, B. (2017). “Updating, Undermining and Perceptual Learning.” *Philosophical Studies* 174(9): 2187-2209.

Moss, S. (2015). “Time Slice Rationality and Acting Under Uncertainty.” In T. Gendler and J. Hawthorne (eds.). *Oxford Studies in Epistemology Volume 5*. Oxford University Press.

Pittard, J. (2019). "Fundamental Disagreements and the Limits of Instrumentalism." *Synthese* 196 (12):5009-5038.

Pryor, J. (2000). “The Skeptic and The Dogmatist.” *Noûs* 34(4):517–549.

Pryor, J. (2013). “Problems for Credulism” in C. Tucker (ed.) *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*. Oxford University Press.

Schoenfield, M. (2015). "A Dilemma for Calibrationism." *Philosophy and Phenomenological Research* 91(2): 425-55

Schoenfield, M. (2016). “Bridging Rationality and Accuracy.” *Journal of Philosophy* 112(2): 633-657.

Schoenfield, M. (2018). “An Accuracy Based Approach to Higher Order Evidence.” *Philosophy and Phenomenological Research* 96(3): 690-715.

Smithies, D. (2022). “The Epistemic Function of Higher Order Evidence” in P. Silva & L. R. G. Oliveira (eds.), *Propositional and Doxastic Justification: New Essays on their Nature and Significance*. Routledge.

Srinivasan, A. (2019). “Genealogy, Epistemology and World-making.” *Proceedings of the Aristotelian Society* 119 (2):127-156

Staffel, J. (2025). *Unfinished Business.* Oxford University Press.

Street, S. (2006). “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127 (1):109-166

Titelbaum, M. (2015). “Rationality’s Fixed Point (Or: In Defense of Right Reason)” in T. Szabo-Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology Volume* 5. Oxford University Press.

Vavova, K. (2018). “Irrelevant Influences.” *Philosophy and Phenomenological Research* 96(1): 134-152.

Weatherson, B. (ms.). “Do Judgments Screen Evidence?”

Weisberg, J. (2015). “Updating, Undermining and Independence.” *British Journal of Philosophy of Science* 66(1):121-159.

White, R. (2009). “On Treating Oneself and Others as Thermometer” *Episteme* 6(3):233-250

White, R. (2010). “You just believe that because…” *Philosophical Perspectives* 24(1):573-615.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press.