ANALIZA
Let. 28, št. 2, str. 241–262, december 2024

ANαliZA

HIGHER-ORDER EVIDENCE IN SCIENCE: SOME PROBLEMATIC CONSEQUENCES OF STEADFASTNESS AND LEVEL-SPLITTING

Accepted
1. 10. 2024

Revised
30. 11. 2024

Published
31. 12. 2024

MARTIN JUSTIN
University of Maribor, Faculty of Arts, Maribor, Slovenia
martin.justin1@um.si

CORRESPONDING AUTHOR
martin.justin1@um.si

Abstract Despite our best efforts, we often fail to act in a perfectly rational manner. Recently, some epistemologists have suggested that we should admit our failings and develop a modest epistemology that would take our fallibility seriously. This includes accounting for the role of evidence of our irrationality, usually called higher-order evidence. It seems intuitive that modest reasoners should take such evidence into account. However, it turns out that incorporating higher-order evidence into a principled theory of what rationality requires is not an easy task. In this paper, I first review the debate about higher-order evidence, describing in detail the puzzle of higher-order evidence and the main positions about it in the literature. Then, I provide two novel examples of higher-order evidence, taken from science. I argue that these examples put pressure on the views that reject the role of higher-order evidence. These views commit themselves to the conclusion that some common scientific practices, such as evaluating evidence in systematic reviews or even running null hypothesis significance tests, are irrational.

Keywords
higher-order evidence,
modest epistemology,
evidence,
rationality,
science

https://doi.org/10.18690/analiza.28.2.241-262.2024

AN$_\alpha$LIZA

ANALIZA
Vol. 28, no. 2, pp. 241–262, December 2024

DOKAZI VIŠJEGA REDA V ZNANOSTI: NEKAJ PROBLEMATIČNIH POSLEDIC ODLOČNOSTI IN RAZDRUŽEVANJA RAVNI

MARTIN JUSTIN
Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
martin.justin1@um.si

DOPISNI AVTOR
martin.justin1@um.si

Sprejeto
1. 10. 2024

Pregledano
30. 11. 2024

Izdano
31. 12. 2024

Izvleček Kljub prizadevanju po nasprotnem pogosto ravnamo iracionalno. V zadnjem času so nekateri epistemologi predlagali, da priznamo svoje napake in poskušali razviti *skromno* epistemologijo, ki bo v zakup vzela dejstva o naši iracionalnosti. Pomemben del takšne epistemologije predstavlja tudi opis vloge dokazil o naši iracionalnosti, t. i. dokazil višjega reda. Intuitivno se zdi, da bi morali pri tvorjenju prepričanj takšna dokazila upoštevati. Vendar pa se izkaže, da vključevanje dokazil višjega reda v teorijo racionalnosti ni lahka naloga. V tem prispevku najprej pregledam razpravo o dokazilih višjega reda, podrobno opišem problem dokazil višjega reda in predstavim glavne odgovore nanj. Nato predstavim dva nova primera dokazil višjega reda, vzeta iz znanosti. Trdim, da ta primera pod vprašaj postavljata stališča, ki zavračajo vlogo dokazov višjega reda. Ta stališča implicirajo, da so nekatere vsakdanje znanstvene prakse, kot sta vrednotenje dokazil v sistematičnih preglednih člankih ali celo izvajanje testov statistične značilnosti, iracionalne.

Ključne besede
dokazila višjega reda, skromna epistemologija, dokazila, racionalnost, znanost

https://doi.org/10.18690/analiza.28.2.241-262.2024

Univerzitetna založba Univerze v Mariboru

1 Introduction

One of the more uncontroversial claims in epistemology is that we should strive to have rational beliefs. Regardless of what we think this entails – confirming to our evidence, forming beliefs reliably or responsibly, conditionalizing on our priors and evidence – being perfectly rational is an ideal. We often fail to achieve it. We fail to apply appropriate rules of rationality, we fail to recognize correct rules of rationality, we make mistakes in our reasoning, we get misleading evidence, etc.

Recently, some epistemologists have suggested that we should admit our failings and develop a modest epistemology that would take our fallibility seriously (Dorst, 2020; Christensen, 2020; DiPaolo, 2019). Central to this project is to provide an account of how to deal with evidence of our rational failings. Such evidence abounds. We know that we reason more poorly when sleep-deprived, hungry, or under the influence of drugs; we are demonstratively bad at reasoning about probabilities and exhibit other cognitive biases.

In epistemological jargon, such evidence is usually referred to as higher-order evidence. There is no single agreed-upon definition of higher-order evidence. But to align our ideas: in contrast to evidence that directly concerns some question we try to answer ("I see that the cup on the table is empty."), higher-order evidence concerns our epistemic performance or evidential situation ("I visited my optometrist yesterday, and it turns out my sight is very unreliable.").

It seems intuitive that modest reasoners should take such evidence into account. However, it turns out that incorporating higher-order evidence into a principled theory of what rationality requires is not an easy task. It can lead us down some troubling paths, such as rejecting established epistemic principles, e.g., conditionalization, or admitting that norms of rationality simply give inherently contradictory advice. Modest epistemology has thus encountered resistance in the form of rejecting any bearing or higher-order evidence on the rationality of our beliefs.

In this paper, I will point to an undesirable consequence of rejecting the import of higher-order evidence that has not yet been discussed in the literature. I will argue that higher-order evidence sceptics commit themselves to the view that some

common scientific practices, such as evaluating evidence in systematic reviews or even running null hypothesis significance tests, are irrational. This will not provide a decisive argument against higher-order evidence scepticism. Rather, it will shift the burden of proof on such scepticism to reject the role of higher-order evidence.

The paper is organized as follows. Section 2 presents the problem of higher-order evidence, while Section 3 overviews its main positions. In Section 4, I present two cases of how higher-order evidence is used in today's scientific practice and spell out the consequences of rejecting these cases. The second part of this section also discusses and rejects two arguments against the admissibility of these cases. Section 5 concludes the paper.

2 Higher-Order Evidence: The Problem

To get the discussion going, let us consider the following case:

Experiment. Max is a PhD student in pharmacology. She is working in a team that develops novel antibiotics. As part of her research, she was tasked with testing a new promising compound, *c,* on some e-coli bacteria. To do this, she prepared two rows of Petri dishes: in one, she treated the bacteria with *c*, and in the other, she treated them with one of the existing antibiotics as control. She was doing this late in the evening, after a 10-hour shift, and under considerable time pressure. The next day, she went to check the Petri dishes and observed that there was no difference in the presence of e-coli between the rows. From this, she concludes that *c* is not better than the existing compound. At lunch, however, she read a newspaper article about a new study that found that in 50 % of the cases, researchers in circumstances Max was experiencing yesterday, messed up their experiments. Let us assume that Max was lucky this time and correctly applied the two compounds but has no way of checking that.

In the **Experiment**, Max received two kinds of evidence[1] that in some way concern her beliefs about the efficiency of the new compound. On the one hand, she has first-order evidence, consisting of the experimental results. On the other hand, she

[1] Note that, strictly speaking, one piece of evidence can have both first- and higher-order bearing. Consequently, evidence cannot be neatly separated into different kinds so we should be careful with such "kinds of evidence" talk. When discussing first- or higher-order evidence, I will thus generally mean "evidence that bears on *p* in a first-/higher-order way".

also has higher-order evidence, consisting of the newspaper article reporting on the study results. The problem arises when we consider the direction in which these two pieces of evidence are pointing. Max's first-order evidence suggests *q*: "The new compound is no more effective than the known ones." Meanwhile, her higher-order evidence throws doubt on *q* by suggesting that this belief was formed in an unreliable way. In other words, Max seems to be in a kind of epistemic bound. She should believe both: (1) that *q* and (2) that it is irrational to believe that *q*. Believing something like "*q,* but it's irrational for me to believe *q*" seems epistemically suspect, to say the least.[2]

There are two distinct but related ways of making the problem of higher-order evidence more precise. One presents it as a conflict between what could be called substantive and structural epistemic principles (this diction is from Whiting (2021); see Lasonen-Aarnio (2014), Worsnip (2018), Horowitz (2022), Ye (2022) for presentation along these lines). The other presents it as a problem of higher-order uncertainty (Dorst, 2024; Henderson, 2022). The upshot of both presentations is very similar – higher-order evidence can wedge a gap between our first- and higher-order doxastic states, which seems problematic – but they can differ in how they carve up the individual positions in the debate. In the remaining of the section, I will present both ways of understanding the problem.

Let's start with the view that the problem of higher-order evidence has to do with a conflict between different kinds of epistemic principles. Consider the following plausible substantive epistemic principle:

> *Evidentialism*[3]: S is justified in believing *p* at t if and only if S's evidence at t on balance supports *p*. (Feldman, 2009)

As we saw in the **Experiment**, Max's total evidence, which consists of both her first- and higher-order evidence, supports two propositions. One is *q*: "The new compound is no more effective than the known ones." The other one is something

[2] In jargon: it's epistemically akratic (Horowitz 2014).

[3] Evidentialist principle like *Evidentialism* is here used only as an example. We could in principle substituted it with any substantive epistemic principle of the form: "S is justified in believing *p* at t iff *p* satisfies some condition C for S at t" (Ye 2022)

like q*: "I am totally unreliable at determining q on the basis of my evidence." Now consider a plausible structural epistemic principle:

> *Bridge*: It is irrational for a person to believe that p and to believe that it is irrational for them to believe that p. (Whiting, 2021; Ye, 2022)

It should be quite clear that having both q and q* violates *Bridge*. But both these beliefs are justified by *Evidentialism*. Thus, it seems that either *Evidentialism* or *Bridge* must give.

Alternatively, the lesson of cases like the **Experiment** can be hashed out in terms of higher-order uncertainty (Dorst, 2024; see also Henderson, 2022, who uses the term conviction). First-order uncertainty refers to a familiar kind of uncertainty about whether some state of affairs obtains. For example, if my roommate promised to vacuum the apartment while I am away during the weekend, I might be uncertain whether the apartment is indeed vacuumed. I know my roommate is usually good at his word, but at the same time, he is a bit sloppy when it comes to cleaning the apartment. So, before I open the apartment door on Sunday afternoon, I might be only 0.7 certain that the apartment is vacuumed. Conversely, higher-order uncertainty refers to the uncertainty about whether my first-order belief or credence is rational. For instance, I might reflect more about my roommate and his cleaning habits and realize that I usually judge him too harshly – I systematically underestimate his zeal for cleaning. Consequently, I might become uncertain whether my credence of 0.7 in the proposition that the apartment is clean is rational, given that I am a biased judge in this case.

More specifically, Dorst (2024) defines higher-order uncertainty as "a unique and precise probability function that encodes the rational degree of belief," or $P(P(q) = t)$, where $P(p) = t$ is my first-order credence in a given proposition. Or, in the form of a principle:

> *Higer-Order Uncertainty*. It is rational to have higher-order uncertainty if and only if there is a proposition q and a threshold t such that you should be unsure whether you should be t-confident of q: $0 < \big(P(P(q)) = t\big) < 1$.

The (intuitive) idea that connects this notion with the debate on higher-order evidence states that higher-order evidence directly bears on this kind of uncertainty. Receiving negative higher-order evidence, e.g., reading the news story in the **Experiment**, increases higher-order uncertainty about the rationality of the first-order belief. Similarly, receiving positive higher-order evidence – if Max would learn, for example, that she is extremely reliable in performing her experiments – can decrease it. The question of higher-order evidence can thus be recast as questions about higher-order uncertainty: Can higher-order uncertainty be rational? And if yes, is there a connection between higher-order uncertainty and first-order beliefs?

As Dorst (2024) shows, this is a fruitful way of understanding this problem. Different positions on the existence and the role of higher-order uncertainty also align well with different answers to the puzzle of higher-order evidence, seen as a conflict between substantive and structural principles of rationality. However, since the notion of higher-order uncertainty is embedded into a specific formal framework, translating between the two ways of carving up the positions sometimes requires a bit more work. For simplicity, I will leave a more detailed discussion of the notion of higher-order uncertainty on the side in the rest of this paper and refer to it only in passing.

3 Higher-Order Evidence: The Positions

To sum up the discussion in the previous section, higher-order evidence presents us with a puzzle of the following pattern (Sliwa and Horowitz 2015):

(1) One's rational beliefs should reflect the bearing of one's (first-order) evidence.
(2) One's rational beliefs should reflect the bearing of one's (higher-order) evidence.
(3) One's rational first- and higher-order doxastic states should not be in tension.

Different views on higher-order evidence differ in how they respond to this puzzle. Steadfast views reject (2) (Tal, 2021; Littlejohn, 2018; Titelbaum, 2015; Kelly, 2005). They argue that higher-order evidence has no rational import, so we should simply ignore it. In higher-order uncertainty talk, they reject the rationality of such

uncertainty: if $P(q) = t$, then $P(P(q) = t) = 1$. Level-splitting views reject (3) (Lasonen-Aarnio, 2014; Worsnip, 2018). They argue that evidence has both first- and higher-order import on our beliefs but that there is no way to reconcile these different impacts. For level-splitters, beliefs like "*p* but it is irrational for me to believe *p*" are rationally permissible, thus (3) must give. In higher-order uncertainty talk, these views accept that any possible level of such uncertainty is permissible. Calibrationist views reject (1) (Ye, 2022; Elga, 2010; Christensen, 2010). They argue that higher-order evidence defeats or brackets the first-order bearing of our evidence. Thus, our first-order beliefs should follow or "calibrate with" our higher-order attitudes. If we are uncertain that our belief that *p* is rational, this belief should be revised to reflect this uncertainty. In the higher-order uncertainty talk, these views accept the possibility of higher-order uncertainty but try to show that there is a systematic connection between it and the first-order attitudes. Finally, Dilemma views accept the puzzle wholeheartedly (Christensen, 2016a; Schoenfield, 2015a, 2015b). They argue that rationality presents us with genuinely incoherent requirements.

While some might be more intuitively appealing than others, all these options are surprising in some way. The rest of this section will briefly present the main motivations and appeal of each of the views and point to some of their problems.

3.1 Steadfast views

Steadfastness is mainly motivated by the idea that these views follow naturally if we take two principles of rationality seriously (see Field 2019 for a concise summary; she, however, does not endorse steadfastness). One of these principles is a form of *Bridge*, already discussed above. Titelbaum (2015) presents the following version:

> *Akratic Principle*: No situation rationally permits any overall state containing both an attitude A and the belief that A is rationally forbidden in one's current situation.

The other principle is a kind of meta-principle. It asserts that principles of rationality apply universally to all agents in all situations.

> *Universality*: Requirements of rationality apply universally to all agents regardless of their situation.

These two principles both seem very intuitive. However, taken together, they have an interesting consequence. They are in tension with the idea that making mistakes about what rationality requires is rationally permissible. Consider this situation. Agent A is in a situation S. A also believes in principle R: "When in S, you are permitted to believe *p*". A (correctly) recognizes that she is in S and that R applies to her (given *Universality*). Thus, she concludes that *p*. Given the *Akratic Principle,* A should not be uncertain whether R is a genuine requirement of rationality. Such uncertainty would imply that she is not permitted to believe *p* in S, which would violate the *Akratic Principle*. In other words, A should be certain that she is correct about what rationality requires of her. That holds for every belief that S justifiably holds. Thus, to justifiably hold any belief, she should think that she never makes any mistakes about what rationality requires of her.

This tension between *the Akratic Principle, Universality,* and the possibility of making rational mistakes motivates Steadfastness. Nevertheless, defenders of Steadfastness still need to provide a story about why we should reject the possibility of making rational mistakes instead of the *Akratic Principle* or *Universality* or rejecting the dilemma altogether.[4] As it turns out, providing this story requires very strong commitments about the access or competence of agents. Titelbaum (2015), for example, argues that agents have a priori insight into what rationality requires of them in a given situation. In his words:

> "Every agent possesses a priori, propositional justification for true beliefs about the requirements of rationality in her current situation. An agent can reflect on her situation and come to recognize facts about what that situation rationally requires. Not only can this reflection justify her in believing those facts; the resulting justification is also empirically indefeasible." (Titelbaum, 2015)

[4] In an interesting turn, Skipper (2021) for example shows that rejecting the possibility of higher-order uncertainty is compatible with Calibrationism, i.e., the view that our beliefs should reflect the higher-order bearing of our evidence.

Littlejohn (2018), on the other hand, argues that the prohibition of mistakes about what rationality requires is implied by us being epistemically competent.

Additionally, Steadfastness also gives unintuitive answers in cases such as the **Experiment.** According to this view, Max should ignore any higher-order bearing of her evidence and remain certain that her belief about the experiment is rational. I will explore this worry that Steadfastness leads to problematic conclusions about such cases in more detail in Section 3.

3.2 Level-Splitting views

Level-Splitting accepts both first- and higher-order import of our evidence but rejects the idea that our doxastic states on these different levels must cohere in the way required by principles such as the *Bridge* or *Akratic Principle*. This view is usually argued for by carefully considering and rejecting other possible answers to the puzzle of higher-order evidence that try to reconcile the three claims.

Let us look in a bit more detail at Lasonen-Aarnio's (2014) presentation of this view. After presenting the above puzzle, she outlines three possible ways in which a theory of rationality could deal with such conflicting recommendations. The first one includes introducing what she calls an "über rule": an overarching epistemic rule that would determine the correct rational response for every possible epistemic circumstance. If such a rule exists, then it could not happen that one would apply it perfectly and at the same time get evidence that one's belief is flawed – the rule would already pre-empt the possibility of receiving such evidence and provide an appropriate response for it. While the existence of such a rule would indeed solve the dilemma, Lasonen-Aarnio is sceptical of the possibility and desirability of such rules. First, the existence of an "über-rule" would imply that all other epistemic rules, if taken to hold universally, are, in fact, wrong. Second, the über-rule would push us towards an undesirable kind of epistemic particularism, where epistemic recommendation would be limited to carefully applying the rule to each individual epistemic situation. Consequently, nothing general could be said about epistemic guidance. In other words, the über-rule cannot play the kind of guidance we expect from epistemic rules.[5]

[5] For a push back against Lasonen-Aarnio's analysis of über-rules, see Kappel (2019). He argues that a version of *Evidentialism* could be considered as a kind of (feasible) über-rule.

Second, Lasonen-Aarnio (2014) considers the view that epistemic rules are hierarchically ordered; thus, in every situation, one of the conflicting rules will override others. Under this picture, an epistemic system would consist of two elements: (1) a set of correct epistemic rules and (2) an ordering relation on these rules, a kind of meta-rule that tells us which rule to follow. This picture might seem promising – in cases like the **Experiment**, such meta-rules could tell us whether, given the specific circumstance, we should follow our first-order or our higher-order evidence. However, since meta-rules are just epistemic rules, we can acquire evidence that we made a mistake in applying them. If this is the case, then we would need another meta-rule to tell us how to resolve this conflict between lower-level meta-rules. As Lasonen-Aarnio convincingly argues, the hierarchical picture of epistemic rules thus either ends up in an infinite regress or posits a kind of über-rule that cannot be defeated.

The third option that Lasonen-Aarnio (2014) presents argues that we could simply admit that situations like the one in the **Experiment** present genuine epistemic dilemmas, where an agent is damned to do something they ought not to. I will discuss Dilemma views in more detail in the next subsection. Here, I will just note that both Lasonen-Aarnio (2013) and Worsnip (2018) argue that such views fail to sufficiently explain why we should take different "oughts" to concern the same, rather than different normative domains. In other words, they admit that first- and higher-order evidence can both have normative force, as dilemma theorists would have it. However, they disagree that these normative forces act in the same domain, namely epistemic rationality.

This finally brings us to the Level-splitting views of higher-order evidence. As already mentioned, these views are similar to the Dilemma views in that they want to preserve both the first- and higher-order bearing of our evidence. But where Dilemma views also accept *Bridge,* the principle that doxastic attitudes on different levels must cohere, Level-Splitting rejects it. As Lasonen-Aarnio puts it, she thinks "that subjects who fail to revise their beliefs in putative cases of defeat are criticizable from an epistemic point of view: they are being unreasonable by failing to take into account evidence about their own cognitive imperfections" (Lasonen-Aarnio 2014). However, she rejects the idea that such epistemic failings are also failings of rationality: "There are epistemic oughts that a subject can violate without thereby being epistemically irrational." Similarly, Wornsnip (2018) argues that the tension

between doxastic states can be rational because *Evidentialism* and *Bridge* operate on different normative domains. *Evidentialism* is a narrow requirement that concerns individual states' (ir)rationality. On the other hand, *Bridge* is a broader requirement that guides reasoning more broadly and does not issue recommendations regarding particular situations.

The main problem with the Level-Splitting Views is that they reject intuitive principles like the *Bridge* or the *Akratic Principle*. Consequently, these views allow sets of beliefs that were above recognized as epistemically suspect, for example: "p but I ought not to believe that p." While not strictly incoherent, many consider such sets problematic enough to reject Level-Splitting on these grounds (Henderson, 2022; Ye, 2022).

3.3 Dilemma Views

Like Level-Splitting, Dilemma views concede that there is no elegant way to accommodate both first- and higher-order evidence into our picture of rationality (Schoenfield, 2015b, 2015a; Christensen, 2016a, 2016b). In contrast to Level-Splitting, they try to preserve *Bridge*. To unpack this difference, consider how the two views evaluate the combined belief "p, but I ought not to believe that p." As explained above, Level-Splitting understands this combined belief as rationally permissible, so agents are not rationally required to revise it. On the other hand, Dilemma views concede that it is irrational; however, agents cannot revise it in a way that would satisfy all the requirements of rationality. In other words, examples such as the **Experiment** put agents in an epistemic bind where every course of action is irrational.

Despite this gloomy outlook, defenders of the Dilemma views argue that some options are less bad than others. In other words, even though an agent who receives misleading higher-order evidence cannot act rationally, there is still one *best* epistemic response available to them. Both Christensen and Schoenfield tie this to the notion of accuracy: if none of the available beliefs is rational, then agents can at least aim for accuracy. Or, as Christensen (2016b) puts it:

> "if [an agent] has very strong anti-reliability evidence about herself, she will see that she is faced with two possibilities: either (a) believing something likely to be inaccurate, or (b) believing something irrational. What should such an agent do? She will not aim for (a), since having high confidence that P is too close to having high confidence that a belief that P is accurate. So, of course, she will aim for (b): she'll aim for accuracy over rationality."

Although the Dilemma views manage somehow to preserve all three desiderata involved in the puzzle of higher-order evidence, they present an unusual picture of rationality. As Ye (2022) points out, the Dilemma theorists might underappreciate the amount of higher-order evidence we acquire in everyday lives. If we try to generalize the view, it thus turns out that we are very frequently engaged in such rational binds and dilemmas.

3.4 Calibrationist views

Calibrationist views argue that, in cases such as the **Experiment,** higher-order evidence overrides or defeats first-order evidence (Ye, 2022). Specifically, in such cases, our beliefs should be revised to reflect the bearing of our higher-order evidence.

These views are primarily motivated by an appeal to intuitions in cases such as **Experiment.** To recall, in **Experiment**, Max, the scientist, received higher-order evidence about her unreliability in conducting experiments. To many involved in this debate, it seems intuitive that this evidence should somehow impact Max's belief about the experiment's results. Building on this intuition, defenders of Calibrationism argue that such higher-order evidence defeats the rational import of our first-order evidence. Specifically, this defeat works by "bracketing" our original first-order evidence. The thought here is that higher-order evidence presents us with evidence that either our epistemic performance or evidential situation is in some way problematic. Since we do not have perfect epistemic access and cannot easily know exactly what has gone wrong, the best policy is to bracket the suspect reasoning and evidence.

On an alternative but connected picture, higher-order defeat acts by evidence disposition: when we get higher-order evidence, we no longer possess the original first-order evidence (González de Prado 2020). As Ye summarizes, it is usually the case that for a proposition to serve as our evidence, we must satisfy some condition with regard to this proposition. For example, De Prado (2020) defends the following condition:

> *Competence:* If an agent is not in a position to competently treat that p as evidence that q, she does not possess that p as evidence that q.

If an agent receives higher-order evidence of unreliable reasoning from *p*, then this agent can be seen as violating *Competence*. Consequently, under the right understanding of evidence, higher-order evidence can be seen as dispossessing agents of evidence.

Regardless of the exact understanding of higher-order defeat, defenders of Calibrationism argue that it forces us to adopt the belief that it would be rational for us to have independently of the bracketed (or dispossessed) first-order evidence. Exactly which belief this is is a contentious matter. Schoenfield (2015a), for example, presents this simple principle:

> *Calibrationism*:[6] If, independently of the first-order reasoning in question, your expected degree of reliability concerning whether *p* at time t is *r*, *r* is the credence that it is rational for you to adopt at t.

Ye (2022) presents a more elaborate picture, which she calls Evidence-Discounting Calibrationism. In contrast to *Calibrationism*, Ye's account does not require us to calibrate our credences. Rather, it states that we should calibrate the degree to which we rely on our first-order evidence in forming these credences. In a principle form:

> *Evidence-Discounting Calibrationism:* If, independently of the first order reasoning in question, your expected degree of reliability concerning whether *p* at time t is *r*, the degree to which you rely on your first-order evidence in forming a credence in *p* that is rational for you to adopt at t should cohere with *r*.

[6] This simple model is also sometimes called the "Thermometer Model" (White, 2009).

I will not go into details about Ye's account here but let us at least clarify what she means by "the degree to which one relies on one's first-order evidence." The idea here is that in forming the new credence, one should aggregate two credences: one that would be rational if one would take first-order evidence into account and the other that would be rational if one would ignore this evidence. The expected reliability then determines how the two credences should be aggregated. If the expected reliability of an agent equals 1, an agent is perfectly reliable. Consequently, all the weight should be given to evidence-based credence. If, on the other hand, the expected reliability of an agent is 0.5, the agent is completely unreliable. Consequently, all the weight should be given to evidence-ignoring credence. Formally: $C_1(H) = xC_0(H|E) + (1-x)C_0(H)$.

Despite their intuitive appeal, the Calibrationist views of higher-order evidence are not uncontroversial. For example, they seem to force us to ignore our evidence (Kelly, 2010; Eder and Brössel, 2019). This is epistemically suspect, especially in cases when we receive misleading higher-order evidence, such as in the **Experiment**. Horowitz (2019) additionally argues that we can know a priori that higher-order evidence will often be misleading, which puts added pressure on Calibrationist views. Additionally, both Christensen (2010) and White (2009) have argued that principles like *Calibrationism* conflict with *Conditionalization*:

> *Conditionalization*: When getting new evidence, one's new credence in a proposition should match one's old credence conditional on the evidence.

Since *Conditionalization* is fundamental to Bayesian epistemology (Lin, 2023) – the dominant framework of dealing with graded beliefs – this is often seen as problematic for Calibrationist views.

4 Higher-Order Evidence and Science

The previous section overviewed the existing positions on the role of higher-order evidence in our epistemic lives. As shown, all positions come with both upsides and problems. In what follows, I will try to put some additional pressure on the views that deny the import of higher-order evidence (claim (2) from the above summary of the puzzle). Specifically, I will present two cases taken from scientific practice in which higher-order evidence is thought to play an important epistemic role. I will

argue that Steadfastness and Level-Splitting commit themselves to the view that scientists, in these cases, act irrationally.

Consider the following case:

Review: Based on anecdotal data and some observational studies, which all show this, Max believes that taking medicine M has a side effect p. This is the correct assessment of where the evidence is pointing. Max then reads a systematic review of evidence about this side effect of M. The review notices that, indeed, all available evidence agrees that the incidence of p is, on average, higher in people who also take M. However, all evidence is of extremely low quality. Specifically, the review notes that because of the low quality of evidence, "the true effect is likely to be substantially different from the estimate of effect," that is, from p.

Anecdotal data and the studies present Max's first-order evidence about p, the side effect of treatment M. On the other hand, she has some evidence about this first-order evidence – the systematic review. This review presents no new evidence about the proposition "p is a side effect of M". Rather, it simply evaluates the evidence that's already available to Max. If we understand higher-order evidence as "evidence which bears on a believer's rational capacities, epistemic performance, or evidential situation" (Horowitz, 2022), the review seems to be a clear example of such evidence.

Situations like the **Review** are not uncommon in science. In 2011, the UN's International Agency for Research on Cancer classified radiofrequency electromagnetic fields – the type of radiation emitted by mobile phones and other electronic devices – as possibly carcinogenic to humans. This decision was made based on some early observational studies. However, in 2024, a large systematic review of all high-quality observational studies, which looked at the available evidence and did not present any new data, concluded that "exposure to RF from mobile phone use likely does not increase the risk of brain cancer" (Karipidis et al., 2024).

Let us look at another case:

Significance. Max was tasked with analysing data from experiments that looked at the efficiency of a new compound, *c**, in reducing the growth of e-coli bacteria. She ran the analysis and found that compared to control, *c*, on average,* reduced the growth of e-coli by 20 %. Based on this result, she concluded that *c** reduces the growth of e-coli". Assume that this is correct: *c** indeed reduces growth. Then she remembered an important thing from her training: she should run a null-hypothesis significance test to check whether the difference in the experimental and control group is statistically significant. In essence, these tests tell us how likely we would get the observed result (in this case, the difference in growth of bacteria in the presence of a compound c*) if the null hypothesis is true – meaning that there is actually no effect (in this case, that there is actually no difference between groups in growth of bacteria and the observed difference is due to chance[7]). Max ran the test and calculated the *p*-value of 0.2. She remembered that publications in her field usually require a value of 0.05 or lower to accept the results as plausible.

As in the **Review** example, Max has two kinds of evidence at play. The experimental data presents her first-order evidence about the effects of *c**. On the other hand, the results of the null hypothesis test represent her higher-order evidence. The additional test presented her with no new evidence directly bearing on the effects of *c**. Rather, it can be understood as evidence about the reliability of her first-order evidence. Like systematic reviews, null hypothesis testing is central to contemporary empirical sciences, with a p-value of 0.05 usually taken as the threshold for "significance" (Nuzzo, 2014; "Points of Significance", 2023).

Review and **Significance** thus both present cases in which higher-order evidence is thought to play an important role in belief formation. However, according to Level-Splitting and Steadfast views of higher-order evidence, taking the results of the systematic review in the **Review** or the significance test in the **Significance** seriously would be a mistake. Both the systematic review and the significance test are misleading since Max's evidence in both cases on balance still supports her original belief. Steadfasters should thus argue that Max should simply ignore her

[7] I am grateful to the anonymous reviewer for helping me clarify the notion of statistical significance here.

higher-order evidence in both cases. Level-Splitters, conversely, could concede that she could become more uncertain that her beliefs in the two cases are rational; nevertheless, they would insist that she can continue to hold these beliefs.

I find these conclusions quite problematic. They imply that scientists who take systematic reviews and p-values seriously act in an irrational manner. That is a strong commitment. It is not enough in itself to reject Steadfastness or Level-Splitting – we could also understand it as an exciting contribution of epistemology to science. However, it puts additional pressure on these views to strengthen their position. There are at least two objections that defenders of Steadfastness or Level-Splitting could make to defend themselves against this charge. Before concluding, I will review these objections and argue that they fail.

4.1 Objection 1: Undercutting rather than Higher-Order Defeat

Defenders of Steadfastness and Level-Splitting could argue that the **Review** and the **Significance** cases are dissimilar to other cases of higher-order evidence and that, consequently, their views do not apply to them. The argument goes as follows. In the **Experiment**, Max received higher-order evidence of her own cognitive failings – she learned about *her own* unreliability. In the **Review** and the **Significance**, on the other hand, she received higher-order evidence directly about *her evidence*. The systematic review and the significance test affected her evidence in a much more direct way than reading a study about her unreliability: while the study suggested that she might have misunderstood her evidence, the review and the significance test showed that what she thought counted as her evidence in the given situation was *actually not evidence at all*.

To put this in more technical terms, defenders of Level-Splitting and Steadfastness could argue that the **Review** and the **Significance** are not examples of defeat by higher-order evidence but examples of a defeat of a much more straightforward kind, that is, of undercutting defeat. In short, d is an undercutting defeat for S's belief that p if and only if d is a reason for S to believe that her reasons for believing p are inadequate (Graham and Lyons, 2021).[8] A classic example of undercutting defeat is learning that a red light is shining on a table that looks red: if sensory

[8] This notion is originally due to Pollock (1986).

information about the red table seems to justify the belief that the table is red, the new information about the red light "undercuts" this justification. Both the results of the systematic review and of the null hypothesis significance test can be seen as basically analogous to such red lighting: if Max's results first seemed to justify her beliefs, the results of the review and the test acted as reasons to believe that this justification is inadequate. Since Level-Splitting and Steadfastness do not reject the role of undercutting defeat, they do not claim that scientists should not take systematic reviews and null hypothesis tests seriously.

This objection points to a real ambiguity in defining higher-order defeat and separating it from other kinds of defeat. The difference between various kinds of defeaters is a complex issue that cannot be resolved here. However, I still think the situations in the **Significance** and the **Review** cases more closely resemble the situation in the **Experiment** than cases of normal rebutting defeat. As we saw above, cases of higher-order evidence are problematic because they, on balance, support both beliefs that "p" and "it would be irrational for me to believe that p". That is also the case in **Significance** and **Review**: in both cases, Max's evidence still supports her first-order beliefs. At the same time, she acquired some evidence that her evidence is unreliable and possibly misleading. Thus, her total evidence in both cases supports her first order belief and some degree of higher-order uncertainty.

But this is not the case in the example with the seemingly red table and the coloured lighting. In that example, one's total evidence stops supporting the belief that "The table is red." Rather, the same sensory information now supports a different belief, perhaps "The table is coloured by the lighting," rather than a combination of beliefs "the table is red" and "it would be irrational for me to believe that it's red". In any case, even if we accept that **Significance** and **Review** are indeed cases in which higher-order evidence acts as a simple undercutting defeater, it still rests on Level-Splitters and Steadfasters to explain why their position does not extend to this kind of defeat.

4.2 Objection 2: Wrong Conclusion

The second objection states that we are drawing the wrong conclusions about **Significance** and **Review**. Rather than understanding these cases as uncontroversial examples of ordinary scientific practice, we should see them as highly controversial.

The fact that Level-Splitting and Steadfastness suggest the same should thus be seen as an interesting upshot of these views.

The fact is that the role of systematic reviews and null hypothesis significance testing in science is controversial. Significance testing is often misunderstood (Wasserstein and Lazar, 2016), and the value of systematic reviews is often overstated (Uttley et al., 2023). However, most critics would shy away from recommending completely abandoning these practices. Quite the opposite, most often they suggest a more stringent and even stricter adherence to these practices (Ritchie 2020). The suggestion that scientists should simply ignore higher-order evidence such as systematic reviews or p-values is, as things stand, an interesting but still highly controversial claim.

5 Conclusion

This paper presented the problem of higher-order evidence, reviewed the main positions in the debate, and provided a new argument against the views that reject the epistemic importance of such evidence. Section 2 presented the problem of higher-order evidence, and Section 3 overviewed the main positions on it. In Section 4, I presented two novel cases of how higher-order evidence, taken from the scientific practice. I showed that at least two positions in the higher-order evidence debate give highly controversial answers in these cases. Although this is not reason enough to reject these views, it puts additional pressure on them to strengthen their position. The second part of Section 4 discussed and rejected two arguments against the admissibility of these cases.

References

Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research, 81* (1), 185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x

Christensen, D. (2016a). Conciliation, Uniqueness and Rational Toxicity. *Noûs 50*(3), 584–603. https://doi.org/10.1111/nous.12077

Christensen, D. (2016b). Disagreement, Drugs, Etc.: From Accuracy to Akrasia. *Episteme, 13*(4), 397–422. https://doi.org/10.1017/EPI.2016.20

Christensen, D. (2020). Akratic (Epistemic) Modesty. *Philosophical Studies*, *178*(7), 2191–2214. https://doi.org/10.1007/S11098-020-01536-6

DiPaolo, J. (2019). Second Best Epistemology: Fallibility and Normativity. *Philosophical Studies, 176*(8), 2043–66. https://doi.org/10.1007/S11098-018-1110-Y/METRICS

Dorst, K. (2020). Evidence: A Guide for the Uncertain. *Philosophy and Phenomenological Research, 100*(3), 586–632. https://doi.org/10.1111/PHPR.12561

Dorst, K. (2024). Higher-Order Evidence. In M. Lasonen-Aarnio and C. Littlejohn (Ed.), *The Routledge Handbook of the Philosophy of Evidence* (pp. 176–194). Routledge.

Eder, A. A. and P. Brössel. (2019). Evidence of Evidence as Higher-Order Evidence. In *Higher-Order Evidence* (pp. 62–83). Oxford University Press. https://doi.org/10.1093/oso/9780198829775.003.0003

Elga, A. (2010). "How to Disagree About How to Disagree." In R. Feldman and T. A. Warfield (Ed.), *Disagreement* (pp. 175–186). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199226078.001.0001

Feldman, R. (2009). Evidentialism, Higher-Order Evidence, and Disagreement. *Episteme, 6*(3), 294–312. https://doi.org/10.3366/E1742360009000720

Field, C. (2019). It's OK to Make Mistakes: Against the Fixed Point Thesis. *Episteme, 16*(2), 175–85. https://doi.org/10.1017/EPI.2017.33

González de Prado, J. (2020). Dispossessing Defeat. *Philosophy and Phenomenological Research, 101*(2), 323–40. https://doi.org/10.1111/phpr.12593

Graham, P. J. J. C. Lyons. (2021). The Structure of Defeat. In *Reasons, Justification, and Defeat* (39–68). Oxford University Press. https://doi.org/10.1093/oso/9780198847205.003.0003

Henderson, L. (2022). Higher-order Evidence and Losing One's Conviction. *Noûs, 56*(3), 513–29. https://doi.org/10.1111/nous.12367

Horowitz, S. (2014). Epistemic Akrasia. *Noûs, 48*(4), 718–44. https://doi.org/10.1111/nous.12026

Horowitz, S. (2019). Predictably Misleading Evidence. In M. Skupper and A. Steglich-Petersen, (Ed.) *Higher-Order Evidence: New Essays* (pp. 105–23). Oxford University Press. https://doi.org/10.1093/OSO/9780198829775.003.0005

Horowitz, S. (2022). Higher-Order Evidence. In E. N. Zalta and U. Nodelman (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2022 Edition)*. The Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/higher-order-evidence/

Kappel, K. (2019). Escaping the Akratic Trilemma. In M. Skupper and A. Steglich-Petersen (Ed.), *Higher-Order Evidence: New Essays* (pp. 124–43). Oxford University Press. https://doi.org/10.1093/oso/9780198829775.003.0006

Karipidis, K, D. Baaken, T. Loney, M. Blettner, C. Brzozek, M. Elwood, C. Narh et al. (2024). The Effect of Exposure to Radiofrequency Fields on Cancer Risk in the General and Working Population: A Systematic Review of Human Observational Studies – Part I: Most Researched Outcomes. *Environment International 191* (September), 108983. https://doi.org/10.1016/J.ENVINT.2024.108983

Kelly, T. (2005). The Epistemic Significance of Disagreement. In T. S. Gendler and J. Hawthorne (Ed.), *Oxford Studies in Epistemology, Volume 1* (pp. 167–196). Oxford University Press.

Kelly, T. 2010. Peer Disagreement and Higher-Order Evidence. In R. Feldman and T. A. Warfield (Ed.), *Disagreement* (111–74). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199226078.003.0007

Lasonen-Aarnio, M. (2013). Disagreement and Evidential Attenuation. *Noûs, 47*(4), 767–94. https://doi.org/10.1111/NOUS.12050

Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research, 88*(2), 314–45. https://doi.org/10.1111/PHPR.12090

Lin, H. (2023). Bayesian Epistemology. In E. N. Zalta and U. Nodelman (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2023 Edition)*. The Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/epistemology-bayesian/

Littlejohn, C. (2018). Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research, 96*(2), 257–72. https://doi.org/10.1111/phpr.12271

Nuzzo, R. (2014). Scientific Method: Statistical Errors. *Nature, 506*(7487), 150–52. https://doi.org/10.1038/506150A.

Points of Significance. (2023). *Nature Human Behaviour, 7*, 293–94. https://doi.org/10.1038/s41562-023-01586-w

Pollock, J. L. (1986). *Contemporary Theories of Knowledge*. Hutchinson .

Ritchie, S. (2020). *Science Fictions. Exposing Fraud, Bias, Negligence and Hype in Science*. The Bodley Head

Schoenfield, M. (2015a). A Dilemma for Calibrationism. *Philosophy and Phenomenological Research, 91*(2), 425–55. https://doi.org/10.1111/phpr.12125

Schoenfield, M. 2015b. Bridging Rationality and Accuracy. *Journal of Philosophy, 112*(12), 633–57. https://doi.org/10.5840/JPHIL20151121242

Skipper, M. (2021). Does Rationality Demand Higher-Order Certainty? *Synthese, 198*(12), 11561–85. https://doi.org/10.1007/s11229-020-02814-w

Sliwa, P and S. Horowitz. 2015. Respecting All the Evidence. *Philosophical Studies, 172*(11), 2835–58. https://doi.org/10.1007/s11098-015-0446-9

Tal, E. (2021). Is Higher-Order Evidence Evidence? *Philosophical Studies, 178*(10), 3157–75. https://doi.org/10.1007/s11098-020-01574-0

Titelbaum, M. G. (2015). Rationality's Fixed Point (or: In Defense of Right Reason). In T. S. Gendler and J. Hawthorne (Ed.), *Oxford Studies in Epistemology, Volume 5* (pp. 253–94). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198722762.003.0009

Uttley, L., D. S. Quintana, P. Montgomery, C. Carroll, M. J. Page, L. Falzon, A. Sutton and D. Moher. (2023). The Problems with Systematic Reviews: A Living Systematic Review. *Journal of Clinical Epidemiology, 156*, 30–41. https://doi.org/10.1016/J.JCLINEPI.2023.01.011

Wasserstein, R. L., and N. A. Lazar. (2016). The ASA Statement on P-Values: Context, Process, and Purpose. *The American Statistician,* 70(2), 129–33. https://doi.org/10.1080/00031305.2016.1154108

White, R. (2009). On Treating Oneself and Others as Thermometers. *Episteme, 6*(3), 233–50. https://doi.org/10.3366/E1742360009000689

Whiting, D. (2021). Higher-Order Evidence. *Analysis, 80*(4), 789–807. https://doi.org/10.1093/ANALYS/ANAA056

Worsnip, A. (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research, 96*(1), 3–44. https://doi.org/10.1111/phpr.12246

Ye, R. (2022). *Higher-Order Evidence and Calibrationism.* Cambridge University Press. https://doi.org/10.1017/9781009127332