

HIGHER-ORDER EVIDENCE AS INFORMATION LOSS

B.A. LEVINSTEIN

ABSTRACT. Higher-order evidence is evidence that you have wrongly or rightly handled other evidence according to rational requirements. According to the *accommodationist* position, you should generally respond to such evidence by adjusting your credences in first-order questions to account for your own potential irrationality. Although accommodationism is intuitive, it recommends some odd behavior, such as violating conditionalization and Good's Theorem. I argue that, on the accommodationist picture, some higher-order evidence is best understood as a kind of information loss akin to forgetting, which results in the same type of epistemic behavior.

0. INTRODUCTION

Sadly, we are all at risk of imperfect reasoning. Consider:

HYPOXIA: You are a pilot flying to Hawaii on a Tuesday morning and are worried whether you have enough gas to make the trip. You know current evidence either indicates there is a 99% chance that you'll make it, or a 1% chance you'll make it. You do some calculations and determine that there's in fact a 99% chance. You then get a message from ground control that informs you that you have hypoxia, which is an acute condition that temporarily impairs cognitive ability. As you know, people who have hypoxia only make correct calculations 50% of the time.

The fact that you have hypoxia is *higher-order evidence*. It provides some reason to think that you are not correctly responding to your first-order evidence. The normative question is what you should think now that there's a high possibility that your reasoning faculties are compromised.

There are two basic positions philosophers have commonly taken. The first I'll call *accommodationism*. According to this view, you should significantly reduce your confidence that you have enough gas. After all, people who have hypoxia tend to err at simple calculations, so it's likely you've messed up and should adjust accordingly.¹

The second is known as *steadfastism*. The idea behind this view is that you should follow an optimal policy for handling your evidence, and that any such policy requires retaining credence .99 that you have enough gas. Suppose there actually is a 99% chance you have enough gas given your evidence. The steadfast claims the right response is then to adopt credence .99 that you have enough gas. To get into the swing of this view, imagine that many pilots were to follow the steadfast's advice in this exact situation. Roughly 99 out of 100

¹Accommodationism is meant to include the position normally referred to as *calibrationism* in the literature. However, calibrationism is committed to the answer that your credence that you have enough gas after learning you're hypoxic should be .5, whereas accommodationism merely requires lowering your credence from .99. See (Schoenfield, 2015).

would in fact have enough gas regardless of whether they were hypoxic. So, the steadfast reasons, the correct response to the total evidence is not budging at all in response to the higher-order evidence.²

Steadfastism strikes many as counter-intuitive, but as we'll see, accommodationism is hard to justify from a theoretical standpoint. In brief, it violates Good's Theorem, violates conditionalization, and exhibits a strange sort of agent-relativity regarding evidence.³ The question for the accommodationist, then, will be how to justify such seemingly strange departures from epistemic orthodoxy. After all, higher-order evidence is a kind of evidence, and the right way to handle it should tend to help us figure out what's true.

This paper presents a novel defense of the accommodationist approach to higher-order evidence that explains these odd features. The basic idea is to assimilate (problematic) higher-order evidence to standard cases of losing information. The supposed theoretical problems plaguing the accommodationist approach to higher-order evidence are also common to forgetting (or gaining information that you have forgotten). Likewise, just as forgetting is an essentially diachronic phenomenon, so too is responding to higher-order evidence. We in turn see what is wrong with the steadfast's view: telling somebody to ignore salient higher-order evidence is the analog of telling somebody to retain credence 1 in propositions she has forgotten.

Here's the plan. Section 1 sets the stage by analyzing Greaves and Wallace's (2006) argument for conditionalization. The next section points out the main apparent theoretical advantages of the steadfast position. Section 3 examines standard cases of forgetting information. Section 4 argues that some higher-order evidence can be modeled as information loss and that it's precisely this type of higher-order evidence that results in the odd behavior advocated by the accommodationist. Section 5 contrasts accommodationism with *calibrationism*. Section 6 generalizes the model in section 4 and briefly addresses some objections. Section 7 wraps up.

1. A SYNCHRONIC ARGUMENT FOR CONDITIONALIZATION

To set up some key issues for later, we will briefly rehearse (and simplify) a familiar argument that one should plan to update by conditionalization, which is adapted from Greaves and Wallace (2006). This rehashing will allow us to see why the argument is synchronic and to develop an important notion of information gain and loss.

Greaves and Wallace's argument is accuracy-based and (at least intended to be) purely epistemic. That is, their argument is meant to appeal to an agent's

²Whether a view counts as an instance of accommodationism or steadfastism, on my terminology, depends only on how it tells you to adjust your credence in first-order questions such as "There is enough gas," not on what it tells you about higher-order questions such as " x is a rational credence to have in the proposition that there is enough gas." Thus, *level-splitting* views that require you to think that you ought to have a credence lower than .99 that there's enough gas while still requiring you to retain credence .99 that there's enough gas count as *steadfast* views. For examples of such views, see (Lasonen-Aarnio, 2014; Coates, 2012).

³For more on the violation of conditionalization, see (Schoenfield, 2018). For more on agent-relativity, see (Kelly, 2005). As far as I know, nobody has previously discussed accommodationism's violation of Good's Theorem.

purely epistemic and not practical goals. Below, we'll also discuss the practical consequences of failing to conditionalize.

1.1. Accuracy. For now, we'll assume that all an epistemic agent cares about is her overall accuracy. The higher her credence in truths and the lower her credence in falsehoods, the better off she is all epistemic things considered.

We'll also assume (though this won't matter much beyond this section) that she measures her accuracy using a *strictly proper scoring rule*. A measure of accuracy is strictly proper if every probability function expects itself to be strictly more accurate than any other credence function.⁴

For illustration, we first provide a paradigmatic example of such a rule. Let Ω be a finite set of possible worlds and \mathcal{F} a set of propositions over those worlds. Given $X \in \mathcal{F}$ and $w \in \Omega$, we let $w(X) = 1$ ($= 0$) if X is true (false) at w . Given a probability c over \mathcal{F} , we then have:

$$\text{BRIER SCORE: } BS(c, w) := 1 - \sum_{X \in \mathcal{F}} (c(X) - w(X))^2$$

As one can verify, for any probability function b and distinct credence function c , $E_b(BS(b)) > E_b(BS(c))$, where $E_b(BS(c))$ denotes the expected Brier Score of c according to b .

Given the assumption that all an epistemic agent cares about is her accuracy, that reasonable measures of accuracy are strictly proper, and the assumption that maximization of expected accuracy is rationally obligatory, we can now see how Greaves and Wallace argue for conditionalization.

1.2. Conditionalization is the Plan with the Highest Expected Accuracy. To argue for conditionalization, we now assume that an agent at t_0 has a probabilistically coherent credence function b_0 over some set of propositions \mathcal{F} generated by state-space Ω . For instance, we can let $\Omega = \{RC, RC\bar{C}, \bar{R}C, \bar{R}\bar{C}\}$, where R means it's raining, and C means it's cold. We let \mathcal{F} in turn be all boolean combinations of states in Ω .

At t_0 , the agent knows she'll learn at t_1 some proposition in a partition \mathcal{E} of Ω . For instance, Dave could know that he'll learn for sure whether it's raining or not at t_1 . In this case, $\mathcal{E} = \{\{RC, RC\bar{C}\}, \{\bar{R}C, \bar{R}\bar{C}\}\}$. The question now is what to plan to do upon learning which element of \mathcal{E} is true.⁵

Now, if all you care about is accuracy, then in one sense the best thing to do is simply assign credence 1 to all truths and 0 to every falsehood. However, that "policy" is not implementable. It can't fully guide an agent since the agent does not have access to which world is actual.

Instead, we need to restrict consideration to epistemic policies that the agent can actually implement given the information she will have at t_1 . Toward this end, Greaves and Wallace restrict policies to functions from elements of \mathcal{E} to probability functions. I.e., a *policy* maps every E in \mathcal{E} to a single probability function.

⁴The propriety assumption is standard in the accuracy literature. Since we're mentioning accuracy here only to illustrate one synchronic argument for conditionalization, we omit an extended discussion of its motivation. See (Pettigrew, 2016).

⁵Schoenfield (2017) generalizes this argument to situations when the agent's possible evidence may not form a partition, and Schoenfield (2018) applies those results to the case of higher-order evidence. However, these complications will not matter for our purposes, so I omit them for simplicity.

So, while Dave would prefer to have different credence functions when RC is true and when $R\bar{C}$ is true, no policy will allow him to do so. If he learns it's raining, he'll have the same credence function in RC -worlds as he will in $R\bar{C}$ worlds, since his evidence doesn't discriminate finely enough to allow him to have different credence functions in those different worlds. However, since he learns for sure whether R , he may have different credence functions in RC worlds and $\bar{R}C$ worlds.

For instance, the following two plans count as actual policies:

- Plan to update from b_0 to $b_0(-|R)$ if you learn R , and to $b_0(-|\neg R)$ if you learn $\neg R$.
- Plan to update from b_0 to a probability function that assigns 1 to RC if you learn R and to one that assigns 1 to $\bar{R}C$ if you learn $\neg R$.

Given this restriction on policies as functions from \mathcal{E} to the set of probability functions, how can we evaluate which policies are preferable to which others? Greaves and Wallace answer that we can evaluate them by their expected accuracy.

That is, for any policy pol , an agent should evaluate how good pol is by calculating its expected accuracy according to her current credence function:

$$E_{b_0}(\text{Acc}(\text{pol})) = \sum_{E \in \mathcal{E}} \sum_{w \in E} b_0(w) \text{Acc}(\text{pol}(E), w)$$

A rational agent will plan to update according to the policy that has the highest expected accuracy. And, as Greaves and Wallace show, conditionalization is the policy that uniquely maximizes expected accuracy.

1.3. Takeaways. Greaves and Wallace's argument concerns what you at t_0 *would like* yourself to do at t_1 . At t_0 you're certain that you'll later learn which $E \in \mathcal{E}$ obtains. And, for each E that you might learn, you have a best plan for what to do: i.e., you'd like to adopt $c_{t_0}(\cdot|E)$ as your new credence function.

This argument, then, only directly shows what you should plan to do at t_0 , not what you should actually do at t_1 . Once t_1 rolls around, there isn't as yet a direct reason to care about what your t_0 -self thought.

We leave open whether and under what conditions Greaves and Wallace's argument has diachronic force. That is, we'll consider it an open question whether accuracy considerations like those advanced here ever do in fact give you reason to set your t_1 -credences equal to your t_0 -conditional credences. Below, we will explore in detail what happens in cases of information loss such as forgetting, both why such phenomena are both essentially diachronic and how they interact with the Greaves and Wallace argument.

2. THE APPEAL OF STEADFASTING

Despite the fact that many find steadfasting counter-intuitive, it's important to see that it in fact has strong apparent theoretical appeal over accommodationism. To see why, let's first note that the Greaves and Wallace argument above seems to lead to steadfasting fairly straightforwardly.

For simplicity, we'll establish the following abbreviations. Let E refer to Carol's first-order evidence on Tuesday. G is the proposition that she has enough gas to make it to Hawaii. H is the proposition that Carol is hypoxic on Tuesday.⁶

Suppose on Monday, Carol knows she will either see E or $\neg E$ on Tuesday and will also learn whether she's hypoxic. Let c_0 represent her Monday credence function.

On Monday, $c_0(G|E) = .99$. So, Carol should plan to update to credence .99 should she learn E if she's certain she's not currently hypoxic. What about $c_0(G|E, H)$? Conditional on E and being hypoxic on Tuesday, Carol should still have credence .99. After all, she knows that whether she has hypoxia at some later time has no impact on whether E actually would indicate that there's a 99% chance she has enough gas.

Now, it looks like Carol's possible evidence meets the Greaves and Wallace criteria. She will either learn EH , $E\bar{H}$, $\bar{E}H$, $\bar{E}\bar{H}$ on Tuesday. Conditionalizing on EH or $E\bar{H}$ leads to credence .99, and as usual, conditionalization is just a function of her evidence, so it's a policy by the Greaves and Wallace definition.⁷

If she follows conditionalization, she'll as a matter of fact be much more accurate on average than if she doesn't. So, conditionalization appears to be the best policy and is in accord with the steadfast recommendation. On the other hand, accommodating H and adopting a credence lower than .99 will conflict with Carol-on-Monday's estimate of what's optimal to do and will also lead to less accuracy on average than conditionalizing. Carol-on-Monday most wants Carol-on-Tuesday to steadfast.

Although we remain neutral on the exact diachronic force of the Greaves and Wallace argument, there is surely an oddity about advising accommodationism here. Carol-on-Monday is rational. By her current lights, conditionalizing is the best course of action. If she foresees that she won't actually conditionalize—i.e., if she currently knows she'll accommodate upon learning H —then she is sure she will in fact do something she does not now approve of. It is uncomfortable for the accommodationist to endorse Carol's Monday self's rationality and her Tuesday-self's behavior, when her Monday-self disapproves of her Tuesday-self.

Second, maintaining credence .99 even after learning H looks best from a third-party perspective. Suppose Bob is a pilot with the same training as Carol and is on the ground. He starts with the same prior as Carol and gets access both to E and the fact that Carol is now hypoxic. What credence does he think is best for Carol to adopt in G ? Again, .99.

On the other hand, accommodationism leads to a weird kind of agent-relativity here. It advises Carol to adopt credence lower than .99 but tells Bob that he should still have credence .99. Usually, two agents with the same information should agree, it seems, whether new information would support or infirm a hypothesis. Accommodationism tells Bob that H is irrelevant for him but relevant for Carol.⁸

⁶Note that H is not indexical, but instead refers just to the timeless claim that Carol is hypoxic at a given time. We return to the issue of indexicality and self-location in section 4.4.

⁷For a similar discussion, see (Schoenfield, 2018).

⁸Whether Bob and Carol really have the same "relevant" total evidence here is tendentious. After all, Carol knows 'I am Carol' and Bob knows 'I am not Carol'. Indeed, some readers may not see this feature as "weird" or theoretically disadvantageous to accommodationism. Nonetheless, we

One final bit of support for steadfastness is not entirely epistemic. Good (1967) shows that if an agent updates by conditionalization, then free evidence is never of negative expected value. To illustrate: imagine you're offered a bet on some proposition X that returns \$1 if X and \$0 otherwise at a price of c cents. You have the option of either deciding whether to take the bet now, or you can wait till you learn whether some alternative proposition Y is true before making your decision. Learning whether Y costs you no money. If you are sure you'll update by conditionalization, then, in expectation you're never worse off if you wait to learn whether Y . Since steadfastness advises updating by conditionalization here, then HOE is always at worst useless in expectation to agents who follow its recommendations.

For accommodationism, though, HOE is sometimes of negative expected utility.⁹ Imagine upon takeoff Carol adopts credence .99 that she has enough gas. She knows at the start of the flight that she is not currently hypoxic. She also knows that in one hour she will be exposed to new atmospheric conditions and will subsequently learn whether she is hypoxic at that time. Let H_{t_1} refer to the (timeless) proposition that she is hypoxic one hour after takeoff. She knows she will maintain her current credence of .99 in G if she doesn't learn H_{t_1} and will update to credence .7 if she does.

Suppose that at takeoff she knows she will be offered a bet for \$.90 that returns \$1 if G and nothing otherwise after she's exposed to the new atmospheric conditions and discovers whether H_{t_1} . Because her current credence in G is .99, she thinks taking the bet will in expectation win her around 8.9 cents. She'd prefer her later self to take the bet regardless of whether she learns H_{t_1} , since conditional on H_{t_1} , she still has credence .99 in G . However, she knows that if she in fact learns H_{t_1} , she'll end up refusing the bet. So, the free information of whether H_{t_1} will cost her in expectation.¹⁰

Thus, we see that steadfastness has strong apparent advantage over accommodationism. If we use the standard Greaves and Wallace framework, we end up with steadfastness. Similarly, if we replace one's current credence as the estimator and use instead either a third-party, we get steadfastness. And, if we think information should never in expectation be harmful when free, we also end up with steadfastness.

Before seeing what's wrong with these brief arguments in support of steadfastness, we turn to a different epistemic phenomenon—that of forgetting.

3. FORGETTING

First consider the following case of forgetting, taken from Talbott (1991):

SPAGHETTI: On March 5, Sam was certain she had spaghetti for dinner on March 5. By April 5, Sam will forget entirely what she had for dinner on March 5.

list it here because it has an analog in the case of forgetting, and because the information loss account helps explain why it occurs. For more on such agent-relativity, see (Kelly, 2005).

⁹For clarity: when I say that accommodationism 'violates' Good's Theorem, I simply mean that free information can be of negative expected value.

¹⁰For an extended discussion of the predictably misleading character of higher-order evidence, see (Horowitz, 2018).

Let S be the claim that Sam had spaghetti for dinner on March 5. A few questions arise:

- (1) What should Sam think on April 5 about S ?
- (2) On March 5, what credence should Sam want her April 5-self to have toward S if she knows she'll forget that S ?

It's hard to say what the answer to the first question is. We'll return to this issue below. However, one thing we know is that the answer is not 1. Since she will have forgotten S by then, her credence should be something less than 1.

The answer to the second question—if Sam only cares about accuracy—is that Sam wants her future self to have credence 1 *even though she knows she'll forget whether S* . To see why, note that since, on March 5, Sam knows S is true, she (correctly) believes her future self will be most accurate if she retains credence 1 in S on April 5 even though she will have forgotten. Indeed, the higher Sam's credence in S will be on April 5, the more accurate she'll be according to her March 5 beliefs.

We can, if we like, apply the Greaves and Wallace framework to show this. Let c_0 be Sam's credence function on March 5. On April 5, Sam may know *some* things she doesn't know on March 5: some element E of a partition \mathcal{E} . For each E in \mathcal{E} , Sam determines the credence function she'd most like to adopt. That credence function is just $c_0(- | E)$. But $c_0(S | E)$ is always 1 for every E .

One thing that's happened in this ordinary case of forgetting is that Sam has lost information. On March 5, there is some set W_0 of epistemically possible worlds. That is, W_0 represents all and only the set of worlds that she cannot eliminate as the actual world. On April 5, there's a new set of epistemically possible worlds W_1 . Although Sam may have some information on April 5 that she didn't on March 5, W_1 is not a (proper or improper) subset of W_0 . In other words, there are worlds in W_1 that Sam can't rule out (i.e., worlds where she had something other than Spaghetti for dinner on March 5) that she could rule out originally.¹¹

So, although Sam really does—on March 5—most want her later self to have credence 1 in S even after she forgets, it would be irrational to do what she previously wanted. Moreover, on March 5, Sam is in a position to foresee that on April 5, she won't actually do what she most wants her later self to do. Nonetheless, her credence function on April 5 and her credence function on March 5 can both be fully rational because W_1 is not a subset of W_0 .

It's often said that Greaves and Wallace's argument establishes a synchronic norm that you should *plan* to conditionalize (Pettigrew, 2016). This phrasing, however, is misleading in the context of forgetting. If Sam knows she'll later forget, she knows any 'plans' she forms on March 5 can't really be faithfully and reliably implemented on April 5. What Sam has on March 5 are impotent desires, not plans properly speaking.

In this sense, forgetting is an *essentially diachronic phenomenon*. What you should do after forgetting is not determined by your synchronic conditional

¹¹There are many different ways one may cash out the notion of *information loss*. Here, we invoke the notion of sets of epistemically possible worlds, but a more general notion identifies information with the negative entropy of one's credence function. When one loses information more generally, the loss causes some distribution or other to have higher total entropy. We avoid the more general notion because the technical machinery is not here worth the expositional cost.

desires before forgetting, regardless of your level of rationality at either time. Your earlier desires cannot bind you, and you should, in fact, do something other than what you would have wanted yourself to do. Your two time-slices come apart.

To set up a parallel for later, we can imagine two different epistemological theories. One says Never Forget. Or, at the least, it tells you to keep assigning credence 1 to propositions you've forgotten. The other, Unsure, says that if you've forgotten a proposition, you should generally not assign credence 1 to it.

Although we aren't yet in a position to argue directly against Steadfasting, we can already see Never Forgetting has some analogous theoretical advantages over Unsure as Steadfasting has over Accommodationism. First, as we've just observed, Never Forgetting is in accord with the plans for conditionalizing argued for by Greaves and Wallace since it maximizes expected accuracy.

Second, once again, Sam's learning that she will forget whether *S* apparently violates Good's Theorem if she figures she'll not retain credence 1. On March 5, she'd take a bet for \$.90 that pays a dollar if *S*. She'll want her future self to take that bet too. But if she knows that she'll later assign credence less than .9 once she's forgotten *S*, then she knows she'll refuse a bet that will guarantee a \$.10 profit.

To develop the parallels between normal information loss and higher-order evidence below, it will be useful to consider a second kind of case in which one is unsure whether one has lost information. Consider the following adapted from Arntzenius (2003):¹²

SHANGRI-LA: You have reached a fork in the road to Shangri-La. The guardians of the tower will flip a coin to determine your path. If it comes up heads, you will travel the Path by the Mountains; if it comes up tails, you will travel the Path by the Sea. Once you reach Shangri-La, if you have traveled the Path by the Sea, the guardians will alter your memory so you have an apparent memory of having traveled the Path by the Mountains. If you travel the Path by the Mountains, they will leave your memory intact. Either way, once in Shangri-La you will seem to remember having traveled the Path by the Mountains. The guardians explain this entire arrangement to you, and you believe their words with certainty.

There are two different cases here: one where you travel the Path by the Sea, and one where you travel the Path by the Mountains. In either case, once you arrive in Shangri-La your cognitive state will be the same.

Upon arrival, you gain information that you have lost information about your actual route. In the case where you actually travel by the Sea, your apparent memory is inaccurate. But in the case where you take the Path by the Mountains, it is veridical. Nonetheless, in neither case should you assign credence 1 to the claim that you've taken the Path by the Mountains. That's because you have evidence that you've lost information. Once you arrive, the world where you took the Path by the Sea is epistemically possible for you, even if it isn't while you're on your way.

¹²This case is actually adapted from Titelbaum (2013), which itself adapted Arntzenius's original case.

So, here, evidence you've lost information is itself loss of information. In the Path by the Mountains case, you start with credence 1 you've taken the Path by the Mountains, but then upon arrival you lose this information and assign credence less than 1. In the Path by the Sea case, you start with credence 1 you've taken the Path by the Sea, but then once again upon arrival you lose that information. Therefore, merely by taking into account evidence of information loss, you do, in fact, lose information because new worlds are epistemically possible.

We can now point to a third parallel between accommodationism and our Unsure policy above. Suppose Sam and Dave on April 5 both have equally clear and distinct apparent memories of eating spaghetti on March 5. If Sam and Dave both learn that Sam's apparent memory may have been caused by something other than her actually eating spaghetti, they should obviously respond differently to this information. Dave can still rely with high confidence on his apparent memory and retain high credence that Sam had spaghetti, whereas Sam should naturally moderate her credence that she had spaghetti.¹³

As we'll see below, both these cases of information loss due to forgetting have analogs in the case of higher-order evidence. Moreover, the same erroneous types of considerations that tell agents to have credence 1 in truths they've forgotten lead to the steadfasting view, while considerations that tell agents to account for their own information loss lead to accommodationism.

4. HIGHER ORDER EVIDENCE AND INFORMATION LOSS

This section has three main goals:

- (1) To distinguish between 'weird' and 'not weird' higher-order evidence.
- (2) To show 'weird' higher-order evidence can fruitfully be modeled as information loss.
- (3) To argue that such information loss accounts for the 'weird' features of HOE on the accommodationist view.

'Weird' here simply means that higher-order evidence sometimes exhibits the features identified above—agents should violate conditionalization, Good's Theorem, and so on when they are under its spell. However, as we'll see, some higher-order evidence is more or less just normal evidence that can be handled as information gain.

4.1. Back to Hypoxia. To get into the swing of the view, let's return to our central case of HYPOXIA. Here, we make some suggestive remarks that analogize cases like HYPOXIA to standard cases of forgetting. Below, we'll develop this idea further.

Let's start with a highly idealized case. Suppose that until Carol receives the evidence *H* on Tuesday, she has always been a perfect bayesian, knows she's been a perfect bayesian, and who has credences over a huge set of propositions. So, she knows she has (until now) updated just by conditionalization.

¹³We remain neutral on what Sam and Dave's *evidence* is in this situation. Some philosophers may think that Dave has as part of his evidence that Sam had spaghetti, while Sam does not, whereas others may regard only the phenomenal aspects of their memories as relevant evidence. For our purposes, nothing essential turns on these thorny questions of what counts as evidence.

On Tuesday, when Carol learns H , she suddenly becomes worried that she may have mishandled E . Supposing she knows she was previously ideal, why does Carol not simply defer to her Monday credences?

More explicitly, let c_0 be Carol's credence function on Monday. Let $M = x$ refer to the proposition that $c_0(G | E, H) = x$. That is, $M = x$ means that Carol on Monday had credence x in G conditional on E and the fact that she's hypoxic on Tuesday. Since Carol is ideal, M actually is equal to .99.

If c_1 refers to her Tuesday credences, it seems that $c_0(G | E, H, M = x)$ should be equal to x . If Carol has a perfect memory still, she should simply recall the true value of M and set her current credence equal to it even after learning her reasoning is impaired.

On any interesting version of this story, the answer has to be that Carol does not have access to what her previous conditional credences were! She not only cannot trust her current reasoning abilities but must also not have infallible access to her earlier credences.

In this version, at least, we see that Carol has undergone information loss. On Monday, no epistemically possible world is one where $M = .01$. But on Tuesday, there are such epistemically possible worlds.¹⁴

Less fully idealized versions of this case are slightly trickier. Suppose that on Monday Carol never actually bothered to form a credence in G conditional on E and H . How can we say she has lost information in this case?

Two responses are available here. First, credences need not ever be occurrent. I may have never previously considered the exact proposition that ticket 7 will win in a fair lottery with 19 tickets, but I am still best modeled as having credence $1/19$ in that proposition.

Second, regardless of how to adjudicate this tricky issue of which credences an agent really has, Carol on Monday (assuming she was then rational) had the relevant information about G conditional on E and H somehow or other embedded in her cognitive system. She was in a position to come to form a conditional credence of .99 were she to work it out.

One might object that before she works it out, her epistemically possible worlds are wide open. In one sense, that's true. However, we can also understand a stricter notion of epistemic possibility here—the worlds that would not be ruled out if she were to work out her conditional credences. She is in a position to eliminate worlds at t_0 that she's not in a position to eliminate at t_1 . In this stricter sense, worlds where $M = .01$ are not possible on Monday. But even after she thinks very hard on Tuesday, worlds where $M = .01$ are epistemically possible. So, again, information loss has occurred. Regardless of how this is modeled, we should understand Carol as having information available to her on Monday that she does not have available on Tuesday.

As we've already seen, the HYPOXIA case is *weird*. If Carol is an accommodationist, she'll violate Good's Theorem and conditionalization, just as she would

¹⁴Instead of appealing to earlier actual selves, we could instead appeal to the agent's hypothetical *ur*prior—the credence the agent began life with before getting any evidence. There are two ways to model information loss in terms of *ur*prior. Either we can suppose the *ur*prior u itself was aware of the relevant norms of reasoning, so $u(G | E, H) = .99$, or we model the *ur*prior as ignorant of certain norms of reasoning, just as Carol is on Tuesday, so that $u(G | E, H) \neq .99$. In the first case, Carol 'loses' information insofar as she does not have access to her *ur*prior. In the second case, Carol loses information relative to her Monday-self, but not relative to her *ur*prior.

in normal cases of forgetting. And, at least on this picture, HYPOXIA is essentially diachronic. Even though her Monday-self would like her Tuesday-self to have credence .99 were she to learn E and H , her time slices come apart. She can't know what her Monday-self wanted once Tuesday rolls around because she has lost information relative to her earlier time-slice.

To complete this initial suggestive comparison, let's now consider a variant of HYPOXIA. In this variant, Carol is again flying at a high altitude on Tuesday and tries to determine whether G on the basis of E . This time, however, she does not learn H for sure. Instead, she learns there's a 50% chance that H (so a 50% chance she's currently impaired), and a 50% chance her reasoning faculties are completely in tact.

In the variant version, the accommodationist should still say Carol ought to have a lower credence in G than .99 after learning she might have hypoxia, regardless of whether she in fact does have hypoxia.

This case is like SHANGRI-LA. There, even if you in fact took the Path by the Mountains, you should have credence less than 1 upon your arrival in Shangri-La. As we saw, the reason in that case was that evidence of information loss is itself information loss. Upon arrival in Shangri-La it becomes epistemically possible that she took the Path by the Sea.

In this version, Carol has once again lost information. She has evidence that she has hypoxia. If learning you have hypoxia is information loss, then so too is learning that you might have hypoxia. Before learning that she might have hypoxia, there were no epistemically possible worlds where $M = .01$, but now there are.

Indeed, as in SHANGRI-LA, such information loss is predictable ahead of time. Carol could, on Monday, know that she'll end up with a 50% chance of impairment on Tuesday. On Tuesday, she would still undergo information loss, and she knows that such a loss will occur on Monday. Similarly, whilst on the Path by the Mountains, you know information loss will occur upon your arrival. Once again, we see a parallel with forgetting.

4.2. Normal Higher-Order Evidence. Before moving forward, we should now distinguish between normal and weird higher-order evidence. We'll fill in some additional assumptions below, but for now, consider:

RAIN: Dave has evidence E and is wondering whether it will rain. He currently assigns credence .6 to rain but then learns that the only rational credence to have on E toward the claim that it will rain is .7.

The information Dave acquired about rationality is higher-order. It tells him about the rational way to handle other evidence. On some straightforward ways of filling in the details of this case, we can think of Dave as initially unsure of what's rational. There are epistemically possible worlds where the only rational credence to have on E toward the proposition it will rain is .1, others where it's .2, etc. Somehow or other, Dave's views about the rational credence to have in rain affect his credence in rain. When he learns what the rational credence to have is, he eliminates all worlds where the rational credence is anything other than .7 and changes his credence in rain accordingly through conditionalization.¹⁵

¹⁵Dave now has different evidence from before, since he knows not just E but also what the rational way to handle E is. So, his new credence in rain may—depending on the right account

Because conditionalization is sufficient to satisfy Good's theorem, Dave would also expect this sort of information about rationality not to be practically harmful. If given the opportunity, he would rather learn the information before placing bets and not after. Finally, a third party observer who also learned what rationality recommended in this case and started with the same prior as Dave would also update in the same way Dave did.

Before we try to model RAIN and HYPOXIA in any kind of unified way, we can already see the difference. In RAIN, after learning what rationality requires, Dave knows strictly more than he did before. He ends up with strictly fewer epistemically possible worlds than he started with, since he's simply eliminated worlds where rationality requires a credence other than .7. In HYPOXIA, Carol's epistemically possible worlds *are not a subset of her possible worlds on Monday*. On the view suggested here, then, it's only when higher-order evidence causes you to *lose information relative to your earlier self* that it should result in 'weird' behavior.¹⁶

4.3. A Toy Model of Higher-Order Evidence. There are many different ways of modeling the core thesis of this paper—that the 'weird' behavior the accommodationist mandates results from higher-order evidence that causes information loss. To cash out the view, I'll suggest one toy framework here that I hope will shed some further light on how such a view may go, though I will flag now that certain details of this model are not central to the information-loss thesis.

First, and perhaps most controversially, I'll assume the existence of a unique maximally rational response to evidence. That is, there's a single maximally rational *ur*prior that an ideally rational agent without any empirical evidence should begin life with. Further, given any body of total evidence *E* and any proposition *X*, there's one maximally rational credence to have toward *X*.

Second, I'll assume a *level-bridging* view. According to level-bridging principles, your credence in a proposition *X* should in some way be harmonious with your credences about what it's rational to believe about *X* on your evidence. For instance, in RAIN, a level-bridger might suggest that once Dave learns that .7 is rational on his evidence, he should have credence .7 that it will rain. In contrast, there are *level-splitting* views, according to which your credences about what's rational need not have much (any?) influence on what your first-order credences should be.¹⁷

I think for accommodationism to get off the ground, some level-bridging principle is required. After all, if you don't think your views about rationality should affect your first-order beliefs, then it's going to be hard to connect higher-order evidence in any systematic way to your first-order views.

Even with these two assumptions, determining exactly how you should bridge your views about rationality with your first-order credences is a difficult and

of how knowledge of rationality works—not be equal to .7. Still, Dave here should update via conditionalization, so I ignore this wrinkle for the moment. See the discussion of RATREF below.

¹⁶Note that an agent can get weird HOE (and lose information) even if she was never fully rational to start with.

¹⁷See footnote 2 for references.

vexed issue. Because I think the exact details are mostly extraneous to the information-loss picture, I'll assume a very simple—and likely wrong—principle:¹⁸

RATREF: Let $R^E(X) = x$ be the proposition that the uniquely rational credence to have in X given total evidence E is x . If b is your credence function and E is your total evidence, then it ought to be the case that for any proposition X , $b(X | R^E(X) = x) = x$.

The idea here is that you should simply defer straightforwardly to whatever is rational. It also entails that—assuming you're probabilistically coherent—your credence in any proposition just is the *expected value* of the rational credence in that proposition. So, even if you don't know what rationality requires, you always set your credence equal to your best estimate of what you should do.

Let's see how this model will help us make sense of the information loss view. In RAIN, as we've already seen, Dave starts out not knowing what $R(\text{rain})$ is, but then he eliminates worlds where it's anything other than .7.

Next, HYPOXIA. What is $R^{E,H}(G)$? It's tempting for the accommodationist to say that the answer depends on whether it's Monday or Tuesday. In fact, the answer is always .99. R simply refers to the rational credence function, which is itself an abstract object. Whether Carol has hypoxia on Tuesday is not relevant to whether there's enough gas, since the rational credence function itself is not in any way cognitively impaired. It is hard to incapacitate abstract objects.

So, on Monday, Carol is in a position to know what $R^{E,H}(G)$ is. But on Tuesday, there are epistemically possible worlds where $R^{E,H}(G)$ takes on values other than .99. She has lost information, and 'weirdness' occurs.

Let's now look at a second variant version of HYPOXIA. Suppose that, as in the initial variant, Carol learns on Tuesday (at t_1) that there is a 50% chance she's hypoxic. Call her evidence at t_1 B . Then, suppose that t_2 , she learns she in fact is not and never was hypoxic. Call her evidence at t_2 C .

As we've already seen, she gets lossy—i.e., weird—HOE at t_1 . On Monday, she is in a position to know that $R^B(G) = .99$, but at t_1 she is not. What about C ? Obviously, relative to her Monday credences, this is not weird evidence. She knows strictly more, since she now knows that she was not hypoxic at t_1 or t_2 and also knows E (her first-order evidence). But it is also not weird relative to her credence function at t_1 . At t_1 , she has access to worlds where:

- (1) She is not hypoxic and her reasoning led to the right answer.
- (2) She is hypoxic and her reasoning led to the right answer.
- (3) She is hypoxic and her reasoning did not lead to the right answer.

Her credence at t_1 accounts for all three possibilities. When she learns C , she eliminates the second and third types of worlds from the list. So, she strictly gains information about what's rational.

Alternatively, we can imagine that at t'_2 , she learns that she in fact is hypoxic. In this case, her new information is lossy relative to her Monday credences, since she's lost information about the requirements of rationality. However, it's not lossy relative to her t_1 credence, since she simply eliminates worlds of the first type. Her t'_2 credence just is her t_1 credence conditional on being hypoxic.

¹⁸See (Christensen, 2010) for discussion of this principle, and (Elga, 2013) and (Dorst, 2019) for reasons to think it's wrong and more sophisticated alternatives.

We now see how weirdness relates to sub-ideal agents. At t_1 , Carol accounts for the fact that she might not be ideal in a certain way—she may have hypoxia. If she learns at t'_2 that she *does* have hypoxia, then such information is not weird relative to t_1 , as it does not destroy information she possessed at t_1 . Indeed, at t_1 , she'd rather wait to find out if she is in fact hypoxic before taking any bets about whether she has enough gas.

On the other hand, the fact that an agent starts out sub-ideal does not preclude him from acquiring new 'weird' higher-order evidence. Suppose, for instance, Charles is at a party and has had one drink at t_1 . He knows even small amounts of alcohol sometimes affect his cognition, but he nonetheless tries to work out a problem from his physics homework and arrives at credence .95 that the answer is 7 meters/second. His credence takes into account the chance that the single drink has impaired his reasoning ability. He knows that in the next few minutes, however, he'll drink five more drinks in quick succession, which will almost surely impair his memory and reasoning. After binge-drinking (at t_2), he reasons again through his physics problem but knows he's too drunk to trust his own judgment.

At t_1 , he'd much rather trust his t_1 -judgment than his future t_2 -judgment about his physics homework. If he had the option at t_1 either to commit then to a bet on whether the answer is indeed 7 m/s or to wait till t_2 to decide whether to commit to the bet, he'd rather make his decision at t_1 . So, even though Charles began potentially sub-ideal, new HOE can still result in information loss.

4.4. Indexicality. One aspect of higher-order evidence we've been ignoring so far is that it is *indexical*. In HYPOXIA, Carol is only affected by H when *it's now Tuesday*. It matters where she is in the world.

We already know that conditionalizing on indexical information is not generally the right thing to do. After all, if you always update just by conditionalization, then once you have credence 1 in a proposition, you'll retain credence 1 forever after. But if I learn that it's now 12:03 PM, then I should not have credence 1 in that claim ten minutes from now.

Some have supposed that it's this indexicality that explains at least some of the 'weird' behavior of the accommodationist view.¹⁹ It's worth examining whether that's true, and whether that casts doubt on the information-loss view.

First, we note that it can't simply be indexicality that explains all the weirdness. If I know that it's currently 12:03, then I would most want myself in ten minutes to be sure that it's 12:13. However, in HYPOXIA, Carol on Monday most wants herself to have credence .99 in G should she learn both E and H on Tuesday. But on Tuesday, the accommodationist recommends a credence lower than .99 to Carol. So, there's discord in the latter case between your earlier wishes and your best future actions, but not in the former case.

Second, not all indexical higher-order information behaves weirdly. Consider:

BLOOD-TYPE: Felicia is highly confident that the European Union will be in tact for another hundred years. She arrived at this view by starting her life with a prior b and updating it by conditionalization. She is sure she has never updated in any other way. She then learns that people with her blood-type tend to start life with a prior that is overly confident

¹⁹See, e.g., (Christensen, 2010).

in long lasting unions between nations. How should Felicia update her views?

It depends on the specifics of the case. If Felicia is sure that $b = R$, then she has nothing to worry about. After all, she's certain her prior was rational, and she's sure she always updated by conditionalization.

On the other hand, if she started life unsure that $b = R$, then she acquires new evidence relevant to what R could be. However, when she learns that b likely assigns higher credence to a claim than R does, she can simply update via conditionalization on this information as usual. That is, she simply gains information about the value of R .

Either way, though, this higher-order evidence is not weird. But it is indexical. It is only relevant to Felicia because *she has the relevant blood-type*. If she were sure she didn't have the relevant blood-type, then it wouldn't matter to her.

It may seem that it's indexicality that accounts for one 'weird' feature flagged above, *viz.*, agent relativity. If Hannah is a third party with a different blood type who initially shared Felicia's optimism about the EU, then learning the same information Felicia did shouldn't necessarily change her opinion. However, it's important to see here that Hannah and Felicia actually have different information. Felicia knows her own prior views and blood type, but she doesn't know about Hannah's. Felicia may still end up with a lower credence than she started with even after learning about Hannah, but such behavior is not due to any essential agent-relativity. Hannah too may lower her credence after learning she started with the same views as people with Felicia's blood-type. Once they each know what the other knows, the accommodationist need not give them different advice about how to respond to their evidence.

Finally, note that cases of forgetting and information loss are, in some ways, essentially indexical to begin with. When you forget information, the information loss is *relative to what you knew at an earlier time*. The essential diachronicity of weird higher-order evidence depends on the temporal locations of different time-slices. So, while higher-order evidence is indeed indexical, it's not the mere indexicality that explains its weirdness, but instead the fact that its part and parcel of the diachronic phenomenon of losing information.

4.5. Plans and Diachronicity. Schoenfield (2018) has recently argued that the steadfast advises the agent to conform to the best plan, whereas the accommodationist is interested in the best plan to make. To see the difference, note that on Monday, Carol wants to *follow* the plan to update by conditionalization, since the best plan in expectation to follow is indeed updating by conditionalization.

However, even if Carol most wants to *follow* the plan of updating by conditionalization, she knows she likely won't be able to if she ends up hypoxic. Instead, a simpler plan that allows her to hedge her credence that she has enough gas in the event of hypoxia may be a better plan to *make*. Compare: perhaps the best plan for the recovering alcoholic to follow is to have, on occasion, one drink at dinner but no more. However, the alcoholic knows she can't actually follow such a plan, so instead makes a plan to abstain from alcohol entirely.

The picture presented here is not in any deep disagreement with Schoenfield. However, as we saw in the case of standard forgetting, information loss is an essentially diachronic phenomenon. One cannot always even *make* plans, however simple, that can later be implemented. Indeed, Carol's Monday plans

are better described as desires, since they cannot be followed faithfully by her later self. What's missing from Schoenfield's account, then, is an understanding of the diachronic element common to higher-order evidence and forgetting.

5. ACCOMMODATIONISM AND CALIBRATIONISM

So far, I've been cagey about exactly what credence agents should assign in cases such as HYPOXIA. All I've said is that accommodationism, in that case, requires assigning a credence less than .99.

One standard version of accommodationism is a view called *calibrationism*, which has more definite prescriptions. According to calibrationism:²⁰

CALIBRATION: If an agent's expected reliability regarding X is r , then that agent's credence in X should be r .

Recall that in the original version of HYPOXIA, Carol knows both that she has hypoxia and that people who have hypoxia make correct calculations only 50% of the time. One is tempted to say that in this case, Carol should expect that she has 50% reliability in determining whether G and in turn should have credence .5.

But as Isaacs (ms) persuasively argues, this view commits the base-rate fallacy. To see why, suppose Carol was not calculating whether she had enough gas to make it to Hawaii, but instead was calculating whether she had enough gas to make it two miles further. In this situation, even though pilots are on average only 50% reliable, Carol should presumably have credence more than .5. More generally, it can't be the case that regardless of what Carol is trying to calculate—however complicated or simple, however outlandish or mundane—she should now have credence .5 in her answer.

What should Carol's credence be instead? I don't have an answer because this problem is simply an instance of the *problem of the priors* that haunts bayesianism (and much of non-bayesian epistemology) generally.

Standardly, bayesian agents start with a prior credence which they (in normal circumstances) update by conditionalization. However, there is no agreed upon solution to the question of what their priors should be.

Carol's Tuesday credence function is 'disconnected' from her Monday credence function because she has lost information. She cannot leverage her earlier credences to determine her new credences. Instead, she has to come up with a good credence function given the information she has. But epistemology currently lacks a method for telling her how to map such information into a credence function.

Of course, one may claim (as we supposed earlier) that there is a single rational credence function. But, in this case, we're supposing Carol does not even know what that is. And we ourselves don't know what it is. So, it's hard to give her specific enough advice here to pin down a single credence.

Note that the problem of the priors arises as well for standard cases of forgetting. In SPAGHETTI, by April, Sam has forgotten what she had for dinner on March 5. What should her new credence be now that she has forgotten? There's no good and implementable solution available. Even if we point to the

²⁰See (Schoenfield, 2015). There may be other ways to interpret CALIBRATION, but those will not clearly recommend a credence of .5 in HYPOXIA.

rational prior updated on Sam's current information, we have no idea what that prior actually tells Sam to think now.

Although this lack of an answer may be some cause for pessimism, there is cause for optimism as well. By assimilating 'weird' types of HOE to cases of forgetting, progress on the latter problem can be turned into progress on the former.

Indeed, there are other potential constraints on credences after either forgetting or HOE that are plausible. In the case of forgetting, Titelbaum (2013) argues for a norm called *generalized conditionalization*. Suppose at t_1 Alice has b_1 as her credence function and has total evidence E_1 , and at t_2 Alice has b_2 and total evidence E_2 . Then, according to generalized conditionalization, it should be the case that $b_1(- | E_2) = b_2(- | E_1)$. The idea here is that, between t_1 and t_2 , Alice may have both gained and lost information that makes her new total evidence E_2 . So, if she's forgotten anything, then b_2 should not simply be b_1 conditional on new evidence since b_1 knows stuff b_2 doesn't. However, Titelbaum argues, if b_2 and b_1 are both brought up to speed on the other's information, then they should agree.

Whether generalized conditionalization can be justified from an accuracy-first perspective is unclear. After all, forgetting is a diachronic phenomenon, and the standard Greaves and Wallace argument only shows agents should *plan* or *want* to update by conditioning. However, it is a plausible constraint on credences at different times even if accuracy-firsters have yet to find a way to nail it down. If it's right for forgetting, it should also be right for HOE.

And, at least in HYPOXIA, it enjoys roughly the same amount of plausibility. On Sunday, Carol knows more about rationality than she does on Monday. However, her Monday credence, conditional on knowing that rationality requires credence .99 in G given E and H should also assign credence .99.

6. GENERALIZATIONS, OBJECTIONS, AND REPLIES

6.1. Rational Urprior. So far, I've developed a model of HOE according to which its odd features are the result of a kind of information loss. Toward this end, I've appealed to a rational urprior that the agent has certain information about, and to which the agent always defers.

Naturally, many (including me) are suspicious of the existence of a rational urprior. Moreover, even those who do believe in it may disagree with RATREF, which spells out the exact way in which agents ought to defer to rationality.

In reply, I claim that appeal to the rational urprior was primarily (if not entirely) for expositional purposes. The rational urprior works nicely (if it exists and if there's a simple deference principle) because the agent can use her estimate of its value to determine what her credence should be at any time.

However, in general, we can often find alternative credences the agent could use for deference purposes instead. For example, in HYPOXIA, Carol would on Tuesday defer to her Monday credences conditional on her current evidence as we've already seen. Her Tuesday credences should just be her best estimate of her Monday conditional credences.

The point is simply that after getting the problematic kind of HOE, the agent has lost some information that she had before. So, she should often be able to find some previous (conditional) credences that she can estimate to determine

what she should currently think now that she has forgotten. The exact details need careful attention on a case-by-case basis, but the general idea stands even without assuming the existence of a rational urprior.

6.2. n^{th} -Order Evidence. A related worry concerns higher orders of evidence. Suppose at first that Brenda subscribes to a principle like RATREF, but then she hears new arguments that cast doubt on its legitimacy and give reasons to support RATREF', which is an alternative deference principle that tells you to defer to rationality in a way that conflicts with RATREF's advice.

Brenda now has third-order evidence. This new evidence provides information about how to handle normal higher-order evidence. What should Brenda do to determine her credence?

I doubt there is a fully general theory of what to do when you don't know what to do when you don't know what to do...in epistemology.²¹ Sometimes, there will still be something to say. E.g., if Brenda thinks that there is some credence function or other she should defer to, then she should use her best estimate of its values to determine her credence. However, there will probably be certain situations in which no single credence function that Brenda can think of will play this role.

Although I don't have answers to these extreme cases, I stand by the claim that when Brenda gains n^{th} level higher-order evidence, she either will simply be updating as usual or she will have undergone some kind of information loss or other (such as losing information about what to do with 2nd-order evidence).

6.3. Incoherence and Logical Uncertainty. Sometimes higher-order evidence plays on our lack of logical and mathematical omniscience. For example, consider:²²

MATH: Anton is an anesthesiologist trying to determine which dosage of pain medication is best for his patient. He applies a standard mathematical formula and calculates the answer to be 3 milligrams. He then is told that he himself may have been exposed to a powerful anesthetic that causes people to be bad at math. How confident should he be that the answer to the math problem is in fact 3 mg?

Here, Anton is unsure about some mathematical claim. However, any agent who's probabilistically coherent (at least in the classical sense) will always have credence 1 in a logical truth and credence 0 in a logical falsehood.

This sort of case presents two difficulties for the theory of higher-order evidence above. The first is that I've appealed to functions like the rational urprior to use for setting one's credence. In particular, I've noted that if Anton will defer to some credence function, then his current credence function should be the expected value of that function. However, if his current credence function is not a probability function, then the notion of *expected value* is undefined.

I reply that his current credence should reflect his best *estimate* of the value of the credence function he should defer to. For probability functions, estimates should nearly always coincide with expected value. But estimates are (plausibly) a more basic type of doxastic attitude (Joyce (1998); Jeffrey (1986)). The closer

²¹See (Sepielli, 2014) for discussion of this issue in the case of ethics.

²²This case is adapted from (Sliwa and Horowitz, 2015).

an estimate is to the true value, the better. So, Anton should try to set his credence to be as close as possible to the one he wants to defer to even if he doesn't have the machinery available to identify his estimate with an expected value.

The second problem concerns the relationship between mathematical uncertainty and information loss. Although Anton in this case may now be highly uncertain of some logical claim due to higher-order evidence, it is not immediately apparent he has lost information.

I think he has. Before worrying about HOE, he had the wherewithal to transform some mathematical information (e.g., axioms) into new mathematical content (e.g., theorems). So, the axioms previously were highly informative to him because he could prove theorems with them. Now that he is worried about HOE, however, he cannot make as much use of the axioms (at least while retaining certainty). So, he has lost information about how to manipulate mathematical symbols. Put differently, before exposure to the drug, were Anton to perform a calculation, all of his epistemically possible worlds are worlds with the right answer to the math problem. After worrying about the HOE, if he were to do the calculation, there are worlds with wrong answers as well.²³

6.4. What Information? Finally, one may worry that there isn't really any information to lose. I've so far appealed to the idea that the agent has forgotten information about some other credence function such as her own previous credences or the rational credences or what have you. However, real life agents rarely think about these things. Agents don't keep meticulous track of what their conditional credences were last Monday, and most people have never given any thought to the one true rational urprior.

This is all true. But for information loss to occur, the agent does not need to consider any of this explicitly. At one point or another, in HYPOXIA, normatively relevant information was encoded in her cognitive system, but on Tuesday, that information has been corrupted. That is, on Monday, somehow, Carol had information available to form the right conditional credences (even if she never did so explicitly), but on Tuesday, the information was erased. Compare: many people fluent English speakers have never given much thought to the explicit rules of English grammar. Nonetheless, these rules must somehow be encoded somehow, and if they suffer a brain injury that damages their ability to speak English, this information is corrupted.

So, regardless of how the information is stored or how much 'direct' access the agent has to it, higher-order evidence can destroy it.

7. CONCLUSION

Higher-order evidence sometimes behaves like normal evidence and sometimes behaves in a weird way. The weirdness—violation of Good's Theorem and conditionalization, agent relativity, and so on—occurs when higher-order

²³Thus, although more work needs to be done to see exactly how well this account can be extended to cover agents who are not even broadly bayesian, there is nothing that requires strict adherence to bayesian norms for the central idea of weird higher-order evidence being a species of information loss to work. For more on worries about bayesianism in the context of higher-order evidence, see (Christensen, 2007).

information causes the agent to lose information she used to possess, even if she did not explicitly realize she had such information.

This diagnosis explains both why accommodationism appears intuitive but theoretically problematic initially and why steadfastism's apparent theoretic advantage is illusory. Steadfastism is the analog of advice that tells agents not to forget and to retain credence 1 even if they do forget.

REFERENCES

- Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy* 100, 356–370.
- Christensen, D. (2007). Does Murphy's Law apply in epistemology? In *Oxford Studies in Epistemology*, Volume 2, pp. 3–31. Oxford University Press.
- Christensen, D. (2010). Higher-order evidence. *Philosophy and Phenomenological Research* 81(1), 185–215.
- Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly* 49, 113–124.
- Dorst, K. (2019). Evidence: A guide for the uncertain. *Philosophy and Phenomenological Research*, Forthcoming.
- Elga, A. (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies* 164, 127–139.
- Good, I. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science* 17, 319–322.
- Greaves, H. and D. Wallace (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind* 115(459), 607–632.
- Horowitz, S. (2018). Predictably misleading evidence. To Appear in *Higher-Order Evidence: New Essays*, Matthias Skipper Rasmussen and Asbjørn Steglich-Petersen, eds. OUP.
- Isaacs, Y. (ms). Two notions of calibrationism.
- Jeffrey, R. C. (1986). Probabilism and induction. *Topoi* 5(1), 51–58.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science* 65(4), 575–603.
- Kelly, T. (2005). The epistemic significance of disagreement. In J. Hawthorne and T. Gendler (Eds.), *Oxford Studies in Epistemology*, Volume 1. Oxford: Oxford University Press.
- Lasonen-Aarnio, M. (2014). Higher order evidence and the limits of defeat. *Philosophy and Phenomenological Research* 88(2), 314–345.
- Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.
- Schoenfield, M. (2015). A dilemma for calibrationism. *Philosophy and Phenomenological Research* 91(2), 425–455.
- Schoenfield, M. (2017). Conditionalization does not (in general) maximize expected accuracy. *Mind* 126(504), 1155–1187.
- Schoenfield, M. (2018). An accuracy based approach to higher order evidence. *Philosophy and Phenomenological Research* 96(3), 690–715.
- Sepielli, A. (2014). What to do when you don't know what to do when you don't know what to do.... *Nous* 48(3), 521–544.
- Sliwa, P. and S. Horowitz (2015). Respecting all the evidence. *Philosophical Studies* 172(11), 2835–2858.

- Talbott, W. (1991). Two principles of bayesian epistemology. *Philosophical Studies* 62, 135-150.
- Titelbaum, M. (2013). *Quitting Certainties: A Bayesian Modeling Framework*. Oxford: Oxford University Press.