Transitional Attitudes and the Unmooring View of Higher-Order Evidence

Julia Staffel
CU Boulder

(penultimate draft – forthcoming in *Noûs*. Please cite final version when available.)

Abstract: This paper proposes a novel answer to the question of what attitude agents should adopt when they receive misleading higher-order evidence that avoids the drawbacks of existing views. The answer builds on the independently motivated observation that there is a difference between attitudes that agents form as conclusions of their reasoning, called *terminal attitudes*, and attitudes that are formed in a transitional manner in the process of reasoning, called *transitional attitudes*. Terminal and transitional attitudes differ both in their descriptive and in their normative properties. When an agent receives higher-order evidence that they might have reasoned incorrectly to a belief or credence towards p, then their attitude towards p is no longer justified as a *terminal* attitude towards p, but it can still be justified as a *transitional* attitude. This view, which I call the *unmooring view*, allows us to capture the rational impact of misleading higher-order evidence in a way that integrates smoothly with a natural picture of epistemic justification and the dynamics of deliberation.

Introduction

Claire and Mallory go to lunch together, and agree to split the bill in half, with a 20% tip. They both calculate their share, with Claire concluding that they each owe $25 and Mallory concluding that they each owe $27. They realize that at least one of them must have made an error, but they don't know who. As it happens, Claire is correct. Hence, Claire has performed correct first-order reasoning to arrive at her answer, but she has credible, yet misleading higher-order evidence that her calculation might be faulty. How should she react? More generally, what is the rational response when one receives misleading higher-order evidence?

Answers to this question can be roughly divided into three categories: Proponents of a steadfast view say that Claire should stick to her answer, because she has reasoned correctly. Conciliationists recommend that Claire should reduce her confidence in her answer, because her higher-order evidence indicates that she likely made a mistake. Level-splitters think that Claire should stick to her belief, but also believe that her belief is not justified since there is a good chance that she made an error. Unfortunately, each of these responses has unsatisfying implications, since they say that agents must either be epistemically akratic or discount part of their evidence.

In this paper I propose a novel answer to the question of what attitude agents should adopt when they receive misleading higher-order evidence. This new view avoids the drawbacks of existing views while preserving their main insights. It builds on the independently motivated observation that there is a difference between attitudes that agents form as conclusions of their

reasoning, called *terminal attitudes*, and attitudes that are formed in a preliminary manner in the process of reasoning, called *transitional attitudes*. Terminal and transitional attitudes differ both in their descriptive and in their normative properties (Staffel 2019b). I argue that when an agent receives higher-order evidence that they might have reasoned incorrectly to a belief or credence towards *p*, then their attitude towards *p* is no longer justified as a *terminal* attitude, but it can still be justified as a *transitional* attitude. This view, which I call the *unmooring view*, allows us to capture the rational impact of misleading higher-order evidence in a way that integrates smoothly with a natural picture of epistemic justification and the dynamics of deliberation.

My discussion will proceed as follows: In section 1, I explain the three leading responses to the problem of misleading higher-order evidence in more detail. In section 2, I take a closer look at the deliberation dynamics of higher-order evidence cases and show how this can help us develop the unmooring view. Section 3 introduces the distinction between transitional and terminal attitudes that serves as the basis for this new view. In section 4, I explain my account of rationally responding to misleading higher-order evidence in detail. I discuss possible alternative views in section 5.

1. Three Ways of Responding to Misleading Higher Order Evidence

The lunch bill example I just introduced provides a good illustration of the structure of misleading higher-order evidence cases. An agent forms a belief or credence about some claim *p* based on a correct deliberation about their first-order evidence, where their first-order evidence is understood to be the evidence they have that directly bears on whether *p* is true. The agent then receives evidence from a credible source suggesting that their reasoning is mistaken or unreliable. This higher-order evidence is misleading, since the agent actually reasoned correctly. The question to be answered is: What is the rational attitude to adopt in response to receiving this evidence?[1] Besides the classic lunch bill case (Christensen 2007), another well-known case is Horowitz's case of the sleepy detective (Horowitz 2014): Police detective Sam correctly determines which suspect stole the jewels, and forms a rational high confidence that Lucy is the thief. When he tells his colleague, she points out that he is very sleep deprived, which has frequently led him to make reasoning errors in the past. Sam realizes that she is right – his past track record calls into question whether he should trust his reasoning about Lucy being the thief.

How should agents respond to receiving misleading higher-order evidence? The conciliationist position (also called "downward push", Smithies 2019) suggests that they should reduce their confidence in their conclusion. Different versions of conciliationism disagree about how much they should reduce their confidence, but they agree that the higher-order evidence exerts rational pressure that invalidates their first-order justification. Accordingly, our agents would not be justified in retaining their original credences in their conclusions. The conciliationist position

[1] Why focus on this particular type of case? It turns out that other instances in which agents receive higher-order evidence are much less puzzling. For example, suppose an agent has reasoned poorly, without noticing their mistake. Standard theories of epistemic justification tend to agree that the agent's conclusion is unjustified. This is unaffected by receiving higher-order evidence that either tells the agent that they reasoned well or that they reasoned poorly. Hence, these kinds of cases don't present much of a puzzle. See for example Pryor (2018) for discussion.

is attractive, because it captures the compelling intuition that it would be irrationally stubborn to stick to one's guns upon receiving credible evidence that one might have made a mistake. Unfortunately, the conciliationist position has some significant drawbacks. It clashes with evidentialism, specifically, with versions of evidentialism on which evidential support is primarily determined by an agent's first-order evidence that bears on the truth of the claim under consideration. All along, Claire's first-order evidence supports that the lunch bill is $25 per person and Sam's first-order evidence supports that Lucy stole the jewels. The fact that they might have reasoned poorly doesn't seem to change this. Yet, according to the conciliationist, they must adopt an attitude that is not supported by their first-order evidence when they receive the misleading higher-order evidence. A further problem that has been raised for conciliationism is that it leads to undue skepticism, since we can almost always find moderately reasonable people who disagree with us on important matters. Lastly, some have argued that conciliationism is self-undermining, because conciliationists seem to be required to abandon confidence in their view when they encounter non-conciliationists.[2] Proponents of conciliationism have offered reasons for why these problems are not decisive, but they remain costs of the view.

Those who consider these problems for conciliationism damning have often adopted a steadfast position instead (also called "upward push", Smithies 2019). This position emphasizes that the agents in our scenarios have in fact evaluated their first-order evidence correctly and have arrived at the attitude supported by their information. Claire and Sam are thus justified in sticking to their initial conclusions and discounting the misleading higher-order evidence. The steadfast view is attractive because it coheres well with (first-order) evidentialism, it's not self-undermining, and it doesn't lead to rampant skepticism. However, it has the obvious drawback that it condones discounting higher-order evidence. This might be fine when the higher-order evidence is in fact misleading, but by stipulation, our agents don't know whether they have reasoned well or made an error. Entirely discounting this information doesn't seem rational. The steadfast view is thus ill-positioned to deliver any guidance to agents. The view seems to encourage them to stubbornly assume they are correct even in cases in which there is good reason to suspect they might not be. Proponents of the steadfast view offer ways to soften the blow of these implications, but this does not alter the fundamental issue that higher-order evidence has little to no impact on this view.[3]

In an attempt to do justice to both the agent's first- and higher-order evidence, some philosophers have suggested level-splitting views. Level-splitters propose that agents who have misleading higher-order evidence should stick to their original conclusions, but that these agents should also believe of themselves that they are unjustified in holding these attitudes. For example, on this view, Sam should be confident that Lucy stole the jewels, and he should also think that his confidence is unjustified. Claire should think that the lunch bill is $25 per person, and also that she

[2] The literature on disagreement and higher-order evidence is vast, so here and below I can only offer a sampling of references to relevant work. Readers may wish to consult the excellent edited volumes by Feldman & Warfield (2010), Christensen and Lackey (2013), and Skipper & Steglich-Petersen (2019) as a starting place. Proponents of versions of conciliationism include Christensen (2007), Elga 2007, Bogardus (2009), and Feldman (2009). For discussions of the criticisms I mention, see e.g. Elga (2010), Carey & Matheson (2013), Fleisher forthcoming, and references therein.

[3] Variations on steadfast views are proposed by e.g. Van Inwagen (1996), Kelly (2005), Smithies (2012) and Titelbaum (2015).

is not justified in holding this belief. Level-splitting is motivated by an attempt to incorporate all of the agent's evidence. The idea is that by holding on to their original conclusion, the agent respects their first-order evidence, and by also thinking that their attitude is unjustified, they respect their higher-order evidence. The problem is that level-splitting endorses epistemic akrasia, i.e., it permits agents to hold attitudes that they themselves think are unjustified. There is an obvious rational tension in both believing *p* and believing that my belief in *p* is unjustified. Interestingly, in some special cases, akrasia might be rational – those are cases in which agents are uncertain about what their evidence is, or where one's evidence predictably leads one to falsity rather than truth.[4] However, in ordinary cases of misleading higher-order evidence like the ones considered here, it is difficult to see how holding epistemically akratic attitudes could be rational.

Much more can be said about how to best formulate and defend each of these views. But my aim here is not to refute existing views, but rather to present an alternative theory that has all of their advantages and none of their drawbacks. More specifically, the unmooring view will be able to capture the steadfaster's idea that what an agent is justified to believe depends on their first-order evidence, it will preserve the conciliationist insight that we should sometimes lower our confidence when receiving higher-order evidence, and it will avoid condoning epistemic akrasia. None of the existing views can capture all three of these claims. In the next section, I will take a closer look at how the three views just discussed interact with theories of epistemic justification, which will help motivate the unmooring view.

2. Two Types of Justification and the Dynamics of Deliberation

In formulating theories of epistemic justification, epistemologists have been following Firth's lead in distinguishing between propositional and doxastic justification (Firth 1978). The basic idea behind these two notions is that we can distinguish between what an agent is justified *to believe*, and what they are justified *in believing*. On evidentialist versions of the distinction, what an agent is justified to believe is thought to depend on what evidence they have. An agent who is justified to believe some claim *p,* or to assign some specific level of credence to *p* need not actually have this attitude; they might have never even considered *p*. By contrast, to be justified *in believing/having a particular credence* in *p*, the agent must actually have the attitude, and they must hold it in a way that is properly based on their evidence. In rough terms, then, we can say that doxastic justification (holding a justified doxastic attitude) requires propositional justification (having justification for that attitude) plus proper basing. Reliabilists, by contrast, see doxastic justification as the primary notion, and propositional justification as the derived notion. Standard reliabilist accounts consider a belief to be prima facie doxastically justified just in case it is arrived at via a reliable process, or, in cases of inferential belief-formation, via a process that is reliable conditional on receiving reliable inputs. A belief is propositionally justified for an agent on this view when there is a reliable process available to the agent by which they could come to have the belief (Goldman 1979). To give a reliabilist account of justified credence instead of justified belief, one can simply replace the target

[4] Proponents of level-splitting views include Williamson (2011), Coates (2012), Hazlett (2012), and Lasonen-Aarnio (2020). For discussion of the (ir)rationality of level-splitting, see e.g. Horowitz (2014), Sliwa & Horowitz (2015).

state that a reliable process must produce. If a reliable process (or ability, on more virtue-oriented accounts) is a process that reliably produces some valuable target state, then this target state need not be a true belief, it can also be a credence. For example, one may identify the target state as the objective probability that is warranted by the agent's grounds (Dunn 2015, Tang 2016, Pettigrew 2018), or as an evidential probability (Comesaña 2018).

In an insightful article, Han Van Wietmarschen (2013) has observed that each of these notions of justification naturally aligns with a view on how to respond to higher-order evidence. The steadfast view seems most plausible when it is interpreted as being about propositional justification (even if this wasn't its originally intended reading). This is because the steadfast view emphasizes that what the first-order evidence supports remains constant, regardless of what higher-order evidence the agent receives. Similarly for reliabilist propositional justification – it remains true that the agents have a reliable process available to them to figure out the answer, regardless of the higher-order evidence they receive. For example, in the lunch bill case, no amount of higher-order evidence can change the mathematical fact that half the bill with a 20% tip comes to $25. Moreover, it is clear from the example that Claire can do the math. Similarly, Sam's first-order evidence will always implicate Lucy as being the jewel thief. He is able to figure this out by reasoning correctly about the evidence.[5] Interpreted in this way, however, the steadfast view no longer tells agents what to do when they receive credible higher-order evidence that they might have made a mistake. The view simply affirms that their propositional justification remains unaffected. The question of what attitude the agent should adopt upon receiving the misleading higher-order evidence remains unanswered.

The conciliatory view, on the other hand, is better interpreted as concerning doxastic justification, according to Van Wietmarschen. Doxastic justification requires that one's attitudes must be properly based on one's evidence, or arrived at via a reliable process. Learning that there is a good chance that one's reasoning is flawed plausibly undermines the justificatory connection between one's attitude and the basis of one's attitude, rendering the attitude doxastically unjustified.[6] Yet, it has been pointed out that that the association between the doxastic notion of justification and the conciliatory view is not as straightforward as one might have thought at first glance.[7] While it seems compelling that higher-order evidence can defeat one's doxastic justification for one's initial conclusion, we have no clear story about what attitude could be doxastically justified for the agent instead. The conciliationist claims that Claire and Sam should reduce their confidence in their favored answers when they receive the misleading higher-order

[5] Notice that a reliabilist might say that Sam's case isn't an interesting case of misleading higher-order evidence, because it's not a case in which he is doxastically justified before receiving the higher-order evidence. This is because he relies on sleep-deprived reasoning. On this way of looking at the cases, the reliabilist should disregard the Sam example and focus on cases like Claire's.

[6] Most recent versions of reliabilism focus on full belief, regardless of whether they are of the process or virtue variety (see for example Lyons 2016, Kelp 2019, Beddor forthcoming). There has been some debate recently about how to best incorporate a treatment of defeat into a reliabilist theory, but the details of this discussion don't affect the arguments I make here (Beddor 2015, Beddor forthcoming). There are also hybrid accounts of justification that combine elements from both reliabilism and evidentialism, but they don't yield different answers to the questions posed here than the views I've already discussed (see e.g. Comesaña 2010, 2018).

[7] Van Wietmarschen himself notices this (p. 418), and Titelbaum points it out in "Return to Reason" (section 5).

evidence. But this reduced confidence can't be doxastically justified according to commonly endorsed views of doxastic justification, on which doxastic justification entails propositional justification. If propositional justification depends on one's first-order evidence or grounds, which is a popular position, only Claire's and Sam's original attitudes can be propositionally justified. And even if we allow propositional justification to depend on both first- and higher-order evidence, it's not clear whether this supports a reduced confidence for Claire and Sam. Claire still has entailing evidence for the $25 answer, even if we add to that the proposition that Mallory got a different answer. And Sam still has strong evidence that points to Lucy as being the thief, even if we add to that the claim that he reasoned when he was tired. Hence, it's hard to see how the reduced confidence recommended by the conciliationist can be propositionally justified, which means that it can't be doxastically justified either.[8] This leaves us in an uncomfortable position: it seems like whatever attitude agents adopt upon receiving misleading higher-order evidence, their attitude won't be doxastically justified. This is precisely the conclusion reached by Silva (2017): He argues that our agents find themselves in a dilemma, in the sense that there is no attitude they can adopt that is doxastically justified. Van Wietmarschen (2013), Palmira (2019) and Titelbaum (2019) also notice this apparent dead end.

The view I propose offers a way out of this quandary that neither ignores the higher-order evidence, nor leaves the agent without any justified attitudes to adopt. The *unmooring view* embraces the suggestion that misleading higher-order evidence defeats doxastic justification, but it also explains which attitudes agents can justifiably adopt upon receiving such evidence. To motivate the view, let's go back to our examples and think about what should happen after Claire and Sam receive the higher-order evidence. Before learning about her disagreement with Mallory, Claire considers the question of how much they each owe for lunch to be settled, and so does Mallory. They both think they have properly calculated their share. But once they realize they disagree, it seems no longer reasonable for them to consider this question settled. Suppose they subsequently drop their confidence in their respective conclusions and assign some confidence to the other person's answer. Clearly, this reduced confidence only be a transitional or intermediary stage in their deliberation. They both know that their evidence warrants full confidence in some answer, but they don't know which one. Hence, both Claire and Mallory will redo their calculations, perhaps this time in a slightly different way, to catch the mistake. They will only consider a potential answer to the question of how much they owe to be a justified conclusion of their reasoning if they arrive at the same answer via a calculation that appears flawless.

We can observe a similar deliberative dynamic in Sam, assuming he is reasonable. He initially considers the question of who stole the jewels to be settled, but then realizes that he shouldn't trust his sleep-deprived deliberations. As a result, he no longer considers the question of

[8] Even David Christensen (2010), a conciliationist, admits that it is hard to see how misleading higher-order evidence could affect propositional justification: "[I]n the case where I'm immune [to the effects of a reason-distorting drug], it is not obvious why my total evidence, after I learn about the drug, does not support my original conclusion just as strongly as it did beforehand. […] The undermining is directed only at the simple deductive reasoning connecting these parameters to my answer. So there is a clear sense in which the facts which are not in doubt – the parameters of the puzzle – leave no room for anything other than my original answer."

who stole the jewels to be settled – perhaps he still thinks that Lucy is the most plausible suspect, but he wouldn't rely on this claim without re-checking his reasoning after taking a nap. Doing so will confirm that he was correct, and he can conclude once again that Lucy stole the jewels.

These more detailed descriptions of the examples bring out features of the agents' attitudes that have been hardly attended to in previous discussions (but see Palmira 2019 for an exception). The attitudes that agents hold towards possible answers to a question can play different roles in reasoning processes, and an attitude's role has implications for its typical descriptive and normative characteristics. The first role an attitude can play is what we standardly call a conclusion of a reasoning process. It can be played by a belief, a credence or a suspension of judgment, depending on the agent's evidence. The initial attitudes that the agents form in our examples before they receive the higher-order evidence, and the conclusions they reach after deliberating for a second time are plausibly classified as playing this role. But what about the attitudes the agents have after they receive the higher-order evidence, but before they get a chance to revisit their reasoning? These attitudes play a different role – they are mere placeholders in an ongoing deliberation – and they differ with respect to some key descriptive and normative properties from attitudes that function as conclusions of reasoning. Distinguishing between the roles attitudes can play in reasoning holds the key to solving the problem of misleading higher-order evidence, or so I claim.

3. Transitional and Terminal Attitudes

3.1 Two Functions for Attitudes in Reasoning

In two recent papers, Julia Staffel has argued that we can better understand the roles doxastic attitudes play in our reasoning if we distinguish between attitudes that function as *transitional attitudes* and attitudes that function as conclusions, or *terminal attitudes* (Staffel 2019b, 2020). We can characterize the two roles attitudes can play by appealing to differences in their typical descriptive and normative properties. The distinction is best introduced by way of an example (adapted from Staffel 2019b).

Detective Fletcher:

Manny has committed a murder and tries to frame Fred for it. Detective Fletcher, upon initially inspecting the evidence, responds as Manny has planned, and becomes 90% confident that Fred committed the murder. However, as she evaluates the evidence more carefully, she discovers incongruencies that ultimately lead her to conclude that Fred was framed, so she reduces her confidence that Fred is the murderer to 2%. She also comes to believe that Fred didn't do it.

This is a case in which an agent attempts to settle a question based on a fixed body of first-order evidence. The evidence in fact exculpates Fred, but due to Manny's skillful framing attempt, this is not immediately obvious. Fully appreciating what the evidence really supports requires careful deliberation, and accordingly Fletcher's credences shift along with her growing insight into the matter. While she ultimately arrives at the conclusion that is justified by the evidence, her earlier credences are significantly different from her final credence that Fred is the killer. These earlier credences reflect her view of the case based on how her reasoning has progressed up to then, but

she does not consider them settled or justified in the same way as the conclusion she reaches when she finishes her deliberation. Yet, even though she thinks that her earlier credences are not her considered and final take on her evidence, she doesn't seem to be epistemically akratic in a way that epistemologists often consider problematic. This is because she considers these transitional attitudes to be mere placeholders, to be revised as her reasoning leads her to a more thorough understanding of her evidence.

According to Staffel's definition, transitional attitudes are attitudes that agents hold towards the answers to specific questions before they have, by their own lights, finished deliberating about how the evidence they currently have bears on these questions. Terminal attitudes are the attitudes that epistemologists are standardly concerned with – beliefs, credences or suspensions that are adopted as conclusions of reasoning. Sometimes a terminal attitude settles a question, like in the Fletcher example, but it doesn't need to. If an agent's evidence leaves open various possible answers to a question, then a rational agent would adopt non-extreme credences in them as terminal attitudes based on their reasoning (and they might subsequently go on to collect more evidence). These are terminal attitudes because the agent takes themselves to have finished their examination of their evidence.

We can think of terminal and transitional attitudes along the lines of different species of the same genus – they share some key commonalities that make them the same type of attitude, but they are also distinguishable insofar as they play somewhat different roles in our mental lives. Both transitional and terminal credences share important features. They are both graded attitudes that range from certainty that p to certainty that $\sim p$, they both encode the agent's confidence in different ways the world might be like, and they are both responsive to evidence and deliberation. However, they differ in that a terminal attitude is taken by the agent to reflect a sufficiently thorough examination of their evidence to answer the question at hand, while a transitional attitude is taken by the agent to reflect a merely preliminary take on their evidence. They also differ with respect to the norms that govern them. These norms arise out of descriptive differences between these attitudes and the different roles they play in our lives.

We can identify at least three differences between typical instances of terminal and transitional attitudes: The first difference has already been mentioned. It concerns the stability or settledness of the attitudes. When we reach a terminal attitude, i.e., an attitude which, by our own lights, is adequately supported by our relevant evidence, we don't typically change this attitude unless we learn new information that bears on it or come to think that our reasoning was problematic. Transitional attitudes, by contrast, are not settled in the same way. They can fluctuate throughout our deliberation even if we haven't acquired new first-order evidence or spotted a potential error. It is in the nature of complex deliberations that our current best estimate of what the final answer might look like often changes throughout the process.

A second descriptive difference concerns the relationship between our attitudes and our actions and assertions. Terminal attitudes tend to be readily available as bases for actions and assertions. Once we have reached a conclusion regarding how our evidence bears on a question of interest, we tend to use this conclusion as a premise in practical reasoning, and we are willing to assert it. By contrast, the use of transitional attitudes for assertion and action is typically much more

circumscribed. If we had asked Detective Fletcher when she first started her investigation who the killer was, she might prefer to say that she hasn't figured it out yet, or she might say that Fred looks like a plausible suspect, but that she has not yet thought about it carefully. It would seem inappropriate for her to say that Fred is the likely killer without any qualification, when she herself does not consider this to be her settled opinion. A similar observation applies to actions. We usually don't act on transitional attitudes, except in special circumstances. For example, we might act on transitional attitudes when doing so has little or no downsides regardless of how the deliberation turns out in the end. Fletcher might ensure that Fred does not have a chance to destroy evidence, or influence potential witnesses, because these actions could be useful, and are low cost, regardless of whether Fred ends up being guilty in the end. Similarly, we might sometimes be forced to act on a transitional attitude when we don't have time to finish deliberating or when we deem a reasoning task too difficult to properly complete. We might encounter both of these scenarios when taking a difficult test. Yet, in general, we prefer to base our actions on terminal, rather than transitional attitudes.

The last descriptive difference worth mentioning pertains to how we update our beliefs and credences when we reason. When we form a terminal attitude, we usually update other attitudes that are logically and probabilistically related in light of it, at least insofar as we are rational. This is not the case, at least not to the same degree, for transitional attitudes. When our reasoning is still in progress, we tend to insulate it to some extent from our remaining web of beliefs. This makes good sense from an efficiency point of view: there is no reason to waste energy on full updates of our belief network based on attitudes whose status is entirely preliminary.

These descriptive differences can ultimately be explained by how transitional and terminal attitudes represent the world to us. Regardless of whether a credence is transitional or terminal, it represents the world to us as having a certain likelihood of being a particular way. When we are still in the process of figuring out what our evidence supports, and our credences are still transitional, we realize that we have not milked our evidence for all it's worth. Since we think at this time that our representations of the world will be improved by further reasoning, it is understandable that we would hold off on using our transitional attitudes as a basis for action, assertion, and updating if we can help it. A result that can be interpreted as supporting this observation is Good's theorem about the value of total evidence. Good (1967) shows that our decisions have higher expected utility if we incorporate all evidence that is available to us (assuming that obtaining and processing the information is cost-free). Hence, on Good's view, a rational agent's terminal attitudes must reflect all of the agent's (first-order) evidence, which makes these attitudes preferable to transitional attitudes as bases for action from an expected utility standpoint.[9]

To be clear, these descriptive differences are intended to characterize paradigmatic cases of terminal and transitional attitudes. This does not rule out the existence of edge cases: a carefully considered transitional attitude that is formed in the final stages of a reasoning process might sometimes be more stable and available for action than a very hastily formed conclusion. Further, how reluctant an agent is to rely on a transitional attitude plausibly depends on how close they

[9] Thanks to Kevin Dorst for pointing this out to me.

think it is to being their final considered opinion, i.e., the conclusion of their reasoning.[10] What ultimately matters for drawing the distinction is the role the attitude plays in reasoning, and even this might sometimes be difficult to determine in particular cases. However, the existence of edge cases does not undermine the usefulness of the distinction. Further, Staffel's account does not claim that agents form transitional attitude every single time they reason. There is no need to form transitional attitudes when an agent can arrive at a conclusion quickly and easily. Rather, transitional attitudes reflect an agent's confidence in different possible answers to a question while they are in the midst of an extended deliberation.

3.2 Distinguishing Norms for Transitional and Terminal Attitudes

The fact that terminal and transitional attitudes play different roles in reasoning has implications for their normative properties. Plausibly, whether an attitude is fitting or successful from a normative point of view depends on whether it properly plays its role, hence, differences in role can make a difference to the justification criteria that apply to it. The normative differences between transitional and terminal attitudes will hold the key to explaining how agents should respond when they receive misleading higher-order evidence. Theories of epistemic justification tend to be designed to apply to terminal attitudes, i.e. attitudes that we adopt once we finish reasoning about a subject. Typically, such theories consider attitudes to be unjustified that don't cohere with the agent's evidence, or ignore part of it, or that are based on superficial reasoning. We can see this clearly in the way the distinction between propositional and doxastic justification is usually spelled out: What is propositionally justified for an agent depends on what is supported by their total evidence. No attitude that is not propositionally justified for an agent can ever be doxastically justified for them. Hence doxastic justification also depends on the agent's total evidence, and additionally, on properly responding to this evidence. The reliabilist take on these notions, while slightly different (see section 2), shares this feature of evidentialist and hybrid views – reliabilist doxastic and propositional justification are designed to apply *to the outputs* of reliable processes or abilities, *not to transitional states* that may be formed during the operation of these processes.

On this view of how epistemic justification works, only Detective Fletcher's terminal attitude, her low credence that Fred is the murderer, is epistemically justified. None of her transitional credences that she forms in the process of reasoning can be doxastically or propositionally justified, since they are not supported by her total evidence. However, it seems quite inappropriate to call these attitudes unjustified in the same way in which it would be unjustified for Fletcher to *conclude* that Fred is guilty. There is a clear sense in which Fletcher's transitional attitudes track the progress in her reasoning in a way that is rational or justified. We can illustrate this by introducing a contrast case in which Fletcher adopts transitional credences that seem much less rational. In this version of the case, Fletcher, upon initially inspecting the

[10] Thanks to an anonymous reviewer for asking me to clarify how transitional and terminal attitudes represent the world, and for suggesting that I include the point about different degrees to which agents can be willing to rely on a transitional attitude.

evidence, realizes that it seems to implicate Fred. (She is still unaware of the incongruencies that reveal the framing attempt.) Yet, instead of forming a high transitional credence that Fred did it, she infers that Fred is very unlikely to be the killer, forming a transitional credence of 2% that he did it. She arrives at this credence by a type of counterinduction, which is an inference strategy on which the agent judges what seems best supported by their evidence at that time, and then infers the opposite of that. This transitional credence clearly seems far less rational than her high transitional credence in the initial version of the case, even though her low transitional credence is not her final considered opinion, and even though this credence will actually turn out to be what her evidence ultimately supports. Counterinduction is simply a bad strategy for interpreting her evidence, regardless of the stage of reasoning she is at. This observation is an instance of a very general phenomenon – in early stages of our deliberation, we often adopt credences that are informed by our initial take on the evidence, and that are very different from the credences we ultimately reach as our conclusions. Not all such credences are equally good or equally rational. However, it would not be justified for us to already adopt the credences we end up settling on, since doing so would be entirely baseless given how far our reasoning has progressed.

This suggests that transitional attitudes can be more or less rational or justified, albeit not in the same way as terminal attitudes. Their function is not to reflect the agent's take on the world based on a suitably complete assessment of their evidence; rather, their function is to reflect the agent's confidence in different ways the world might be based on the agent's reasoning up to that point. Thus, in order to be rational or justified, a transitional attitude needs to properly capture the agent's take on a particular question at that stage of her reasoning, but this does not require that the attitude is based on a reasoning process that is thorough enough to warrant terminating the deliberation. However, rational transitional attitudes must plausibly be part of an adequate deliberative strategy and reflect the agent's insight into their evidence at the relevant stage of reasoning in order to be at least a good preliminary answer to the question the agent is aiming to settle. Staffel (2020) offers a schematic definition of what makes a transitional attitude rational for an agent:

> A transitional attitude *d* is rationally held by an agent as an answer to some question *q* at some time *t*
> just in case
> (I) the agent is using a permissible cognitive process to settle the question *q*, and
> (II) at *t*, *d* is suitably attuned to both (a) the evidence the agent has considered up to *t*, and (b) the manner in which the evidence has been considered or processed, and
> (III) *d* is properly based on the evidence the agent has considered up to *t*, and (ii) the manner in which the evidence has been considered or processed.

This definition is not fully specific, as there can be different ways of filling in the notions of a 'permissible cognitive process', an attitude being 'suitably attuned' to evidence and a 'manner of processing,' and the basing condition in (III). Still, we can see that it vindicates Fletcher's transitional attitudes as rational: her early high confidence that Fred did it is formed as part of an

extended, thorough deliberation about her evidence, so condition (I) is met. Condition (II) is also met, because early on, Fletcher has reviewed all of her evidence, but only grasps how the pieces are connected in a superficial manner. This superficial take on the evidence seems to point towards Fred as being the killer, hence, her high confidence that he did it qualifies as being suitably attuned to the manner in which the evidence has been processed. Condition (III) is also satisfied, as Fletcher properly bases her confidence on her reasoning. The definition also captures why it would be less rational for Fletcher to form a low transitional credence that Fred did it upon first inspecting the evidence: her low credence is not suitably attuned to the evidence she has considered and her understanding of it, because she uses an impermissible cognitive process, i.e., a counterinductive reasoning strategy, to arrive at it.

It is easy to see how this definition of rationality for transitional attitudes differs from standard accounts of justification for terminal attitudes. The main difference lies in condition (II): Transitional attitudes can be rational even if they are only based on part of the agent's evidence, and they can also be rational if they are based on a superficial or otherwise incomplete or flawed interpretation of the evidence (as long as this interpretation is part of a larger permissible reasoning process that allows for mistakes to be corrected). Yet, they still have to be guided by the agent's insight into their evidence at that stage of reasoning in order to give the agent a good *preliminary* idea of what the world is like. Justified terminal attitudes, by contrast, need to be based on a correct interpretation of all of the agent's first-order evidence in order to deliver the most accurate representation of the world that is available to the agent at that time (at least if we assume a somewhat demanding standard of justification for terminal attitudes).[11]

One might worry at this point that introducing two separate standards of rationality for transitional and terminal attitudes is simply too high a price to pay to deal with the problem of higher-order evidence. We should not embrace such a proposal unless we have exhausted alternative options that rely on a single standard of rationality. I will discuss this objection and alternative options after I explain how the current proposal solves the problem of higher-order

[11] How exactly the justificatory standards for terminal attitudes should be spelled out depends on how demanding one thinks epistemic rationality is. We can generally say that it is rational for an agent to terminate a deliberation and adopt the resulting attitude as a conclusion when, by their own lights, their reasoning process has been suitably thorough and responsive to their evidence for the purposes at hand. But what does it mean for a reasoning process to be suitably thorough and evidence-responsive for the purposes at hand? On stricter, more idealized views of rationality, a conclusion of reasoning is rational if the agent reaches it by applying normatively correct reasoning to their total relevant evidence.

Proponents of non-ideal standards of rationality might adopt a more permissive view of what makes terminal attitudes rational, allowing, for example, that terminal attitudes can be rational that are arrived at via some suitable heuristic. On such a view, certain heuristics will count as permissible cognitive processes according to condition (I) of the definition of a rational transitional attitude, and certain conclusions will count as rational that might be judged irrational if we applied stricter normative standards. Which heuristics an agent can rationally use is typically thought to depend on the specific features of the situation, such as the stakes and the agent's resources.

For the cases discussed in this paper, it doesn't matter which standards we apply, since both stricter and more permissive views of rationality will arguably deliver the same verdicts about our examples. But there will be differences between the views' verdicts if we consider cases in which agents use heuristics in their reasoning to make it more efficient, or cases in which the difficulty of a reasoning task exceeds an agent's cognitive capacities. My arguments for the need to apply different standards of rationality to terminal and transitional attitudes apply regardless of the strictness of the rational norms we adopt, but I won't spell them out in further detail here.

evidence. I will argue that single-standard views are not satisfactory. To proceed with my argument, the claims from this section I need to rely on are (i) that standard views of epistemic justification are best construed as applying to terminal attitudes, and (ii) that transitional attitudes can also be more or less rational, but not according to the same conditions as terminal attitudes. With these claims in hand, I can now put together the *unmooring view* of how to respond to misleading higher-order evidence.

4. The Unmooring View of How to Respond to Misleading Higher-Order Evidence

We would like an account of what attitude agents should adopt upon receiving misleading higher-order evidence that lets them respect all of their evidence without being akratic. I propose that we can achieve this if we adopt a more fine-grained categorization of the doxastic attitudes that agents adopt in the cases under consideration, utilizing Staffel's distinction between transitional and terminal attitudes. This distinction is independently motivated by thinking about how reasoners move through complex deliberations, and we can use it to explain what happens when agents receive misleading higher-order evidence.

When the agents in our examples first finish their deliberations, they form terminal attitudes, i.e. attitudes that they endorse as the conclusions of their deliberation. In Claire's case, this is the belief that the lunch costs $25 per person, and in Sam's case, it's a high credence that Lucy stole the jewels. According to standard (non-skeptical) views of epistemic justification, their attitudes are doxastically justified, because they are supported by their evidence and have been arrived at through proper reasoning.[12] Next, our agents receive credible but misleading higher-order evidence that they made a mistake. As other philosophers have argued, receiving information of this kind serves as a doxastic defeater. This diagnosis strikes me as very plausible, and it has been skillfully defended by various philosophers. For example, in a recent article on the basing relation, Ram Neta gives a compelling explanation of why doxastic justification is undermined by credible, misleading higher-order evidence. He argues that having a properly based attitude requires the agent to represent their exercise of the reasoning disposition by which they have arrived at this attitude as being one that provides doxastic justification. Once an agent receives credible evidence that their exercise of this disposition might be defective, they can no longer represent their exercise of this disposition as providing doxastic justification, and hence one of the necessary conditions for holding a properly based attitude is no longer met (see Neta 2019 for details). Paul Silva (2017) also defends this view and offers a variety of possible explanations of why this is the case. My view is compatible with different explanations of how exactly misleading higher-order evidence leads to doxastic defeat, so readers are invited to supplement whichever explanation they find compelling.[13]

---

[12] Though see my comment in footnote 5 that some reliabilists might not want to count Sam as justified, and footnote 11 about alternative standards of justification.

[13] Smithies (2019) also offers an account of why misleading higher-order evidence defeats doxastic justification. However, as Titelbaum (2019) rightly observes, Smithies' account ultimately doesn't give the higher-order evidence enough efficacy to be compelling. Another interesting account of the distinctiveness of higher-order defeat is given by DiPaolo (2018), who thinks that higher-order defeaters are state-given rather than object-given reasons. I think his arguments are compatible with interpreting these defeaters as doxastic defeaters.

As mentioned above, this invites the question of what attitude, if any, agents can be doxastically justified in having once they've received the misleading higher-order evidence. Silva (2017) argues that our agents find themselves in a dilemma, in the sense that there is no attitude they can adopt that is doxastically justified. This naturally follows if one accepts the doxastic defeat line of reasoning – since the propositional justification is unchanged, there is only one candidate for a doxastically justified attitude, and it has been ruled out. Titelbaum (2019) concurs.

There is a sense in which I agree with this – agents can't have a doxastically justified *terminal* attitude right upon receiving misleading higher-order evidence. But this does not mean that they can't have a justified *transitional* attitude. Once an agent receives credible higher-order evidence that calls the quality of their reasoning into question, this defeats their doxastic justification, and thereby unmoors their rationally held *terminal* attitude. The agent thus returns to holding *transitional* attitudes towards the various candidate answers to the question at hand. These transitional attitudes have the function of providing preliminary, constantly updated representations of the world in light of the agent's evolving understanding of their evidence as they deliberate, and their rationality or justification is evaluated according to how well they play this role. This provides a way out of the seeming impasse identified by Silva and others, as we no longer have to accept that there is *no* doxastically justified attitude an agent can adopt upon receiving higher-order evidence. Since justificatory standards for transitional attitudes differ from those for terminal attitudes, we can now identify justified transitional attitudes the agent can adopt without having to endorse unattractive modifications to views of justification for terminal attitudes.

A salient question at this point is whether agents should merely reclassify their attitudes from terminal to transitional, or whether they should also drop their transitional confidence in their initial conclusions at this point. The key factor in answering this question is in what way the higher-order evidence calls into question the reasoning that led to the agent's original answer. Which parts of the original reasoning process are deemed defective and in what way determines how far the agent is set back in the process of deliberating about the question at hand, and it also determines which transitional attitudes are appropriate to adopt in light of this. The agent might either downgrade their trust in specific reasoning steps they have executed, or they might dismiss those steps altogether, depending on the higher-order evidence they receive.[14]

In Sam's case, it seems reasonable for him to be very worried about having made an error, which would lead him to erase or at least downgrade his trust into all the steps of his reasoning that he thinks might be affected by his sleep deprivation. This doesn't need to be all of his reasoning; perhaps he remembers ruling out some suspects when he wasn't so tired from the long investigation yet. If he thinks, for instance, that his reasoning was fine up until he was left with two suspects, and that his sleep-deprived reasoning was no more reliable than guessing, then he should assign equal

[14] Some conciliationists have made efforts to formulate independence principles that make precise how an agent should bracket their initial reasoning when arriving at a conciliatory credence upon receiving higher-order evidence. My position differs in crucial ways from standard conciliationism, but discussions of independence will likely be very relevant for determining what transitional attitudes agents should adopt. For a recent paper on this, see Christensen (2019).

credence to each suspect being the thief until he can think about what his evidence supports with a rested mind.

We can make similar observations about Claire. Depending on the specifics of the situation, she might reasonably be more or less worried that she is the one who made a mistake, which would lead her to reduce her confidence in her answer to different degrees.[15] In standard presentations of the case, Claire and Mallory are said to be equally good at math, and neither of their answers is implausible on its face. On this description, it is rational for Claire to give equal confidence to their answers, and perhaps to give a small amount of credence to the option that they are both mistaken. However, if we vary the parameters of the case, a different credence distribution might be rational. For example, if Claire has good reasons to trust her own reasoning more than Mallory's, or Mallory's answer seems a bit too high or too low to be plausible, she might reduce her confidence in her own answer, but not to the point where she gives equal weight to Mallory's answer.

Regardless of the exact way we fill in the details here, Claire should consider the credences she forms after learning about the disagreement to be merely transitional. She obviously shouldn't consider a 50/50 or 60/40 credence to be a potentially appropriate conclusion of her reasoning – after all she knows that it's a simple math problem with a correct answer. She knows she can arrive at a definitive answer if she just spends a bit more time re-doing her calculations. So, just like in Sam's case, her reduced confidence can only be justified as a transitional but not as a terminal attitude.

The claim that the attitudes that are appropriate for Sam and Claire to adopt are transitional is not only supported by the fact that they are not appropriate conclusions for their respective deliberations in light of their first-order evidence. It is also plausible in light of other descriptive features these attitudes exhibit, assuming Sam and Claire behave in ways we tend to consider reasonable. I explained in the previous section that transitional attitudes are different from terminal attitudes in that their availability as bases for assertion and action is much more limited. Sam, upon realizing how sleep-deprived he was, would no longer tell people that Lucy did it, or move to arrest her. Rather, he might say that he was suspecting Lucy but had to double-check his inferences to be sure, or he might even prefer to just say he was still thinking about it. Similarly, Claire (assuming she dropped her confidence) would not assert 'It's 50/50 that the bill is $25', rather, she would likely prefer to say she was still checking her math, or that she came up with the answer of $25, but that she might be mistaken because her friend had a different result.[16]

---

[15] The problem of finding a principle that can reconcile one's higher-order uncertainty about what the rational credences are with one's first order credences has turned out to be difficult to solve. As Dorst (2020) shows, the standard reflection principle has counterexamples. He proposes a principle called *Trust* that constrains which combinations of higher-order and first-order credences an agent can rationally adopt. The formal structure of the principle could be interpreted as constraining a person's transitional credences in light of their higher-order uncertainty.

Another approach to integrating first- and higher order evidence that could inform what transitional credences agents should adopt is the calibrationist view (e.g. Sliwa and Horowitz 2015), but see Isaacs (forthcoming) for critical discussion.

[16] It's an interesting question what agents should do when they receive (misleading) higher-order evidence that their reasoning was flawed, but they no longer care about the right answer (perhaps the restaurant offered to comp Claire and Mallory's lunch, making it irrelevant how much they would each owe). In this case, it is a waste of time to redeliberate. One thing we could say is that the agents in these cases are rational to leave in place their transitional

More generally, I submit that the unmooring view delivers two attractive payoffs: First, it enables a better analysis of what happens when agents receive higher-order evidence, which helps us better understand how to divide up the logical space of possible views. Second, it allows us to adopt a view of propositional and doxastic terminal justification that preserves the popular idea that your beliefs and credences should be supported by your first-order evidence in order to be justified, but that avoids the problematic dogmatism of steadfast views. I will explain each of these in turn.

The first payoff concerns the analysis of cases in which agents receive (misleading) higher-order evidence. Before the distinction between transitional and terminal attitudes is introduced, the question that philosophers were trying to answer was simply "which attitude should people adopt when they receive misleading higher-order evidence?" It was expected that a specific rational attitude (a specific credence or perhaps a binary attitude like belief, disbelief, or suspension) could be identified. However, the epistemological tools and concepts that people were bringing to this question were implicitly geared towards (what we can now call) terminal attitudes. While some philosophers acknowledged that the agents should keep deliberating (especially Palmira 2019), the significance of this insight for the quest to identify a rational attitude the agent could adopt before redeliberating was not appreciated. This created significant tensions in trying to solve the problem, since people tried to apply concepts intended for terminal attitudes to a situation in which a rational agent could have no such attitude.

We saw this earlier: Steadfasters claim that justified attitudes must be supported by the agent's first-order evidence. While this seems like a reasonable demand for terminal attitudes, this leaves no space for agents to distribute their credences in a more cautious way when they have not fully evaluated their evidence yet, or when the quality of their reasoning is called into question. If we combine the steadfast demand that justified credences must be supported by the agent's first-order evidence with standard views of propositional and doxastic justification, it can easily seem like there is *no* attitude the agent can justifiedly adopt upon receiving misleading higher-order evidence, as I explained above. The conciliationist view, by contrast, allows for higher-order evidence to interact with first-order evidence in such a way that agents may alter their credences upon receiving higher-order evidence. However, this view allows agents to assign significant amounts of confidence to claims whose falsity is entailed by the agents' evidence. The lunch bill case is an example of this. This strikes many philosophers as an unacceptable view of justification for terminal attitudes, and even conciliationists admit that it is a cost of their view (e.g., Christensen 2010). Lacking the distinction between terminal and transitional attitudes, existing versions of

attitudes, since they don't need terminal attitudes for action or further reasoning. We could also say that they adopt "anti-interrogative" attitudes, which is a category recently introduced by Lord (2020). For redeliberation to be a rational response to receiving higher-order evidence that one's initial reasoning was flawed, it is not sufficient that the higher-order evidence is credible. The agent must also have a continued interest in answering the question at hand. If the question no longer matters to the agent, it is better to leave the matter unsettled and avoid wasting cognitive resources. Furthermore, as mentioned in Sam's case, there is no point in redeliberating if the agent is not in a position to improve on their previous reasoning in some way. Sam shouldn't redeliberate while he is still tired, doing so would be a waste of cognitive resources. Thanks to an anonymous reviewer for encouraging me to mention this.

conciliationism don't limit their proposed standards of justification to only transitional attitudes, for which they are far more plausible than for terminal attitudes.

Once we introduce the distinction between transitional and terminal attitudes, we gain access to conceptual resources that help us see that we don't need to treat the attitudes agents adopt immediately upon receiving higher-order evidence in the exact same way as we treat terminal attitudes, or conclusions of reasoning. Further, since this distinction can be developed independently by thinking about how complex deliberation works in general, we gain a larger context that helps us avoid ad hoc solutions to the problem of misleading higher-order evidence.

This leads me to the second payoff of the unmooring view, which concerns how we should think of doxastic and propositional justification for terminal attitudes. Staffel (2020) proposes a schematic account of what makes *transitional* attitudes rational that is responsive to the quality and stage of one's reasoning, which means that it can incorporate the role of higher-order evidence and accommodate conciliationist intuitions. I will not develop this view here, but consider how it frees up space for adopting a view of the rationality of terminal attitudes that aligns with long-established accounts, such as reliabilism or evidentialism. These views emphasize that correct/reliable reasoning from one's first-order evidence is what really matters for justification, but they struggle to find a role for higher-order evidence. The resulting steadfast position about how to respond to higher-order evidence thus seems ultimately too dogmatic.

Yet, with the distinction between transitional and terminal attitudes in hand, we can now explain why these traditional, first-order-evidence-based views actually succeed in giving plausible standards of justification for terminal attitudes. We can accept that an agent's first-order evidence determines which terminal attitude towards a claim is *propositionally* justified for them. Further, for an agent to have a *doxastically* justified terminal attitude, it must be properly based on this first-order evidence (or be reached via a suitable process of reasoning). Given the role we have specified for higher-order evidence, which gives the agent information about the quality of their reasoning, the view must also make explicit that having a doxastically justified terminal attitude is incompatible with the presence of doxastic defeaters.[17] If the agent receives credible higher-order evidence that calls the quality of their reasoning into question, they can't have a justified terminal attitude until the relevant parts of their reasoning have been checked or replaced. This view retains the steadfast intuition that it can't be rational to arrive at conclusions that are at odds with one's first order evidence, and it avoids the dogmatic implications of the original steadfast view by giving higher-order evidence the power of doxastic defeat.

This view strikes me as very attractive, since it relies on intuitions about the cases under consideration that previously seemed difficult to accommodate by a single theory. I also want to emphasize that readers can accept that transitional and terminal attitudes play distinct roles and have distinct standards of justification, and that higher-order evidence can be a doxastic defeater, while having some freedom about how to fill in these accounts of justification. The basic idea is that transitional attitudes are justified when they properly reflect the agent's insight into what the

[17] Depending on one's epistemological leanings, one might adopt the stronger view that the agent must explicitly rule out such defeaters, or the weaker view that the agent doesn't possess any such defeaters.

world is like at intermediate stages of reasoning, which means that they are sensitive to higher-order evidence and need not be supported by the agent's total first-order evidence. By contrast, terminal attitudes must be justified in light of suitable completed reasoning processes about the agent's first-order evidence, and must not be subject to higher-order defeat. This leaves room for adopting, for example, an evidentialist or reliabilist theory of the justification of terminal attitudes, and for adopting ideal or non-ideal standards of justification. A proponent of an ideal conception of rationality might propose Bayesian norms of rationality as the standards by which we should evaluate terminal attitudes. But it's also possible to adopt standards of non-ideal, or bounded rationality as the correct standards of rationality for terminal attitudes. Due to their computational limitations, human reasoners cannot fully comply with norms of Bayesian rationality, which has led to ongoing debates in philosophy and cognitive science about more computationally feasible norms (see e.g. Dallmann 2017, Icard 2018, Staffel 2019a, Staffel 2020, Vul et al. 2014, Weisberg 2020). Even if we adopt more computationally feasible norms of rationality for terminal attitudes, the question of which standards are appropriate for judging transitional attitudes remains live. I hope that the question of what the right justification norms for transitional and terminal attitudes are will generate a lively debate in future discussions of higher-order evidence and deliberation.

5. Alternative Views

The unmooring view proposes to solve the problem of misleading higher-order evidence by appealing to different functional roles played by doxastic attitudes, with different rational norms applying to the attitudes depending on their role. However, endorsing separate standards of rationality for attitudes depending on whether they play the role of transitional or terminal attitudes is a highly revisionary proposal, so it is worth asking whether a view that endorses a single standard of rationality for evaluating terminal and transitional attitudes could do the job.[18] I will initially rely on the Detective Fletcher case from section 3 to examine the viability of a single-standard view, and then return to considering higher-order evidence cases.

In order to evaluate single-standard proposals, it will be useful to explicitly state some desiderata that a successful account must meet based on our previous discussion. First, such an account should be able to deliver the intuitively correct rationality verdicts about an agent's evolving credences in cases of complex deliberation. It should be able to distinguish between more and less rational attitudes to have at different stages of reasoning processes. For example, the account should count Detective Fletcher's initial high credence and concluding low credence that Fred is the murderer as rational, as explained above. It should also deliver the verdict that if Fletcher adopted alternative credences at any of those stages of reasoning instead, those would be less rational. Second, the account should work with our stipulation that the agent's first-order evidence remains the same throughout their deliberation (otherwise, we'd have to completely redescribe what happens in these cases, which would be undesirably revisionary in itself). Third, the view should not count agents as akratic just because they are in the process of reasoning.

[18] The following discussion draws on suggestion from an anonymous reviewer regarding how a single-standard view might work, as well as arguments from Staffel (2020).

Fletcher's evolving credences don't instantiate epistemic akrasia as it is intuitively understood, and it is a strike against a view of rationality if it says they do. Fourth, the view should allow us to draw a connection between an attitude's rationality and its appropriateness for its role in reasoning. Standards of rationality are related to an attitude's functional role, hence, we should be able to explain why rational terminal and rational transitional attitudes aren't equally available to play all the same roles. For example, we should be able to explain why Fletcher's rational concluding credence would be a suitable basis for asserting that Fred is innocent, but her transitional high credence in his guilt would not be a suitable basis for asserting that he is guilty.

Any view that judges Fletcher's earlier and later credences that Fred is the murderer by a single standard of rationality needs to appeal to some difference between them in order to be able to evaluate both of them as rational. Appealing to changes in the first-order evidence possessed by the agent is ruled out by our second desideratum. Hence, any plausible view must appeal to changes in the agent's higher-order evidence and/or their insight into their first-order evidence in order to explain why different attitudes are rational for them at different stages of their reasoning. This type of view meets the first desideratum. It can explain why Fletcher's initial high confidence and her later low confidence that Fred did it can be rational – her understanding of her first-order evidence and her higher-order evidence about the thoroughness of her reasoning clearly differ at those two stages of her reasoning.[19]

However, this view has trouble explaining why her earlier attitudes cannot play all the same roles as her later ones, and hence runs into trouble with our fourth desideratum. If two attitudes are of the same type (credences in this case), and they are rational according to the same standard (which they must be according to this view, to satisfy the first desideratum), then they should be equally available for reasoning, action, assertion, and whatever other roles they might play.[20] But only Fletcher's later credence is an appropriate basis for asserting who the killer is, deciding who to arrest, etc. We need an explanation for why this is.

An anonymous reviewer suggests that we could appeal to an agent's second-order attitudes for this purpose. On this proposal, what roles a first-order credence can play depends on the agent's second-order credence that their first-order attitude is the correct credence to adopt in light of their evidence. The higher the agent's second-order credence that their first-order credence is their best effort to represent the world, the more available the first-order credence is as a basis for action, assertion, and further reasoning. On this view, the limited availability of transitional credences for playing various roles is explained by the agent's low confidence that this attitude is their best take on their evidence, whereas the more general availability of terminal attitudes for playing various roles is explained by the agent's high confidence that their conclusion is an accurate interpretation of their information.

---

[19] This type of change is also appealed to by Staffel's account of what makes transitional attitudes rational at different stages of reasoning.

[20] More precisely: if two attitudes are of the same type, and they are rational according to the same standard *to the same degree*, then they should be equally available for playing various functional roles. Two attitudes could both clear the threshold for rationality, yet one could clear it by a greater margin, which could make it more available for guiding action, assertion, etc. Yet, it doesn't seem like appealing to a difference in the margin by which Fletcher's attitudes clear the threshold for rationality is going to help the proponent of the single standard view here.

This view delivers the correct verdict if we apply it to Fletcher's concluding low confidence that Fred is the murderer. She has a rational high credence that this attitude is the most accurate interpretation of her evidence, which explains why she terminates her deliberation and is ready to base assertions and actions on her first-order attitude. What about her initial high credence that Fred did it, when she hasn't yet noticed the subtle inconsistencies that point towards the framing attempt? Recall that the single-standard view under consideration is designed to meet the first constraint. Hence, it counts Fletcher's initial high credence that Fred did it as a rational response to her total evidence at that stage of her deliberation. What should her second-order credence be that this first-order attitude is her best take on her evidence? The first option is that her second-order credence should be high, which makes sense given that the single-standard view counts her credence as a correct interpretation of her total evidence at that time. But if her second-order credence is high, we can't explain why her first-order credence is not available as a basis for action, assertion, or reasoning in the same way as her concluding low first-order credence. Further, we can't explain why she doesn't terminate her deliberation. The second option is that her second-order credence that her first-order credence is her best take on her evidence should be low. This seems reasonable considering that she has not finished deliberating by her own lights, and it would explain why her first-order credence's availability as a basis for action and assertion is limited. However, now we're running into a conflict with the third desideratum. This view makes Fletcher look akratic: her first-order credence is rational, because it is the correct way of responding to her evidence at that point in time. Yet, if she is rational, she should also have a low second-order credence that her first-order credence is a correct interpretation of her evidence. This problem does not arise for the two-standards view: a transitional attitude can be rational for an agent, and the agent can think that it is rational insofar as it represents their best current, *non-final* take on their evidence, while also thinking that the attitude is unlikely to be their best *final* take on their evidence. There is nothing akratic about being in *this* state of mind, and it makes sense of why the agent is unwilling to act on their credences. [21]

Let's take stock here: we've tried to account for the rationality of transitional and terminal attitudes in reasoning by appealing to a single standard of rationality that is responsive to first-order and higher-order evidence, in combination with an agent's second-order credences that her first-order credences constitute accurate interpretations of her evidence. We now see that this view

[21] As an anonymous reviewer points out, whether Fletcher's attitudes are akratic depends on our understanding of akrasia. On a strong formulation of akrasia, "an epistemically akratic agent believes something she believes is unsupported by her evidence" (Horowitz 2014). On a slightly weaker formulation, epistemic akrasia is "a mismatch between the doxastic states one is in, on the one hand, and one's beliefs (or states of confidence) about what doxastic state it would be epistemically rational for one to be in" (Lasonen-Aarnio 2020). If Fletcher is merely uncertain whether her 90% confidence is the rational response to her evidence, but doesn't have high confidence or a belief it is not, her attitudes are not akratic on the stronger understanding of akrasia (though they might be on the weaker understanding). However, this can't save the single-standard view, because there are other, similar cases in which it can't avoid attributing akratic attitudes to the agent. In the case of Claire, discussed again below, Claire *knows* that her 50% credence that the lunch bill is $25 is not the credence supported by her first-order evidence. She knows that the rational credence is either 0 or 1, and that she can figure out which it is. Hence, Claire's combination of first- and second-order credences would clearly count as akratic, even on a strong understanding of akrasia. This means that the single-standard view can't meet all the desiderata in every case of interest.

faces a dilemma when applied to transitional credences: On the first horn, the view can't explain why transitional attitudes are less available for playing various roles in reasoning, and why agents don't terminate deliberation upon forming a rational transitional attitude. On the second horn, it can't explain why agents who are in the process of reasoning aren't epistemically akratic.

A proponent of the single-standard view might try to escape this dilemma by pointing out that terminal and transitional credences are just different types of credences. But if we go in for this proposal, we are now back to a view on which each type of attitude is judged by their own norms, since this is the only way to capture all of our desiderata. While it is of course still up for debate whether Staffel's (2020) account is the best way of spelling out these two standards, my argument demonstrates that we need some type of two-standard account that functions along these lines.[22]

This line of argument shows that a single-standard view can't meet all of our desiderata regarding the nature of transitional and terminal attitudes. It will be instructive to apply this view to the lunch bill case to show that it fares no better in the context of solving the problem of higher-order evidence. In the lunch bill case, we want to say that Claire's initial high confidence that the bill is $25 each is rational, and that once she learns that Mallory disagrees, the rational response is to lower her confidence and assign equal credence to the $25 and $27 answers. The single-standard view we have been considering can accommodate this, since it says that when Claire gains higher-order evidence, this changes which credences are rational for her. What should Claire's second-order credence be that her 1/2 credence in each of the salient answers is the correct response to her evidence? Here, we run into the same issue as above: If we say Claire's second-order credence should be high, because she is in fact correctly responding to her evidence, we can't explain why she should reopen deliberation, and why she doesn't use those credences for action and assertion. Alternatively, we might say that Claire's second-order credences that her evenly divided first-order credences are her best take on her evidence should be *zero*, because she knows that she has entailing evidence for the right answer. But if we say this, then the view seems to say that Claire is (and should be) epistemically akratic, which is an undesirable verdict. Further, the view can't explain how Claire can evaluate her evenly divided credences to be more rational than some alternative credences, for example ones that give twice as much confidence to Mallory's answer than to her own. Her second-order credence that these unevenly divided credences are correct should also be zero, hence, she would evaluate them from her point of view as no better or worse than the evenly split credences that are intuitively most rational. This is problematic in light of the first desideratum above. As I explained above, the unmooring view that appeals to terminal and transitional attitudes with distinct standards of rationality can avoid this dilemma, hence, it fares better than the single-standard view regardless of whether we evaluate it in contexts of complex deliberation or with regard to its ability to solve the problem of higher-order evidence.

Another proposal for solving the problem of higher-order evidence has been put forth by Wedgwood (2019), who appeals to norms of non-ideal rationality to explain what attitudes agents

[22] Staffel (2020) also gives a further, though less decisive, argument for introducing distinct standards of justification for transitional and terminal attitudes, which has to do with the appeal of preserving an intuitive notion of propositional justification. Interested readers are invited to consult her paper on this.

should adopt when receiving misleading higher-order evidence.[23] Wedgwood endorses a steadfast view for ideal rationality, according to which misleading higher-order evidence has no impact on what is ideally propositionally and doxastically rational for an agent. However, he thinks that these standards are too demanding for normal, non-ideal recipients of misleading higher-order evidence. He argues that "in these cases, if you were thinking as rationally as is realistically possible for you to do, you would respond to acquiring this higher-order evidence by raising your credence" in whichever claim your higher-order evidence misleadingly suggests as being supported. Wedgwood's view initially seems to be a two-standards view, so one might think that it fares better than the single-standard view just discussed. However, on closer inspection, this is not the case. Wedgwood thinks that the types of cases we've been discussing only arise for non-ideal agents, who should be judged according to a single standard of non-ideal rationality. This view is thus essentially a version of the single-standard view I discussed above, which means that it runs into the same dilemma.

What about alternative accounts that appeal to distinctive attitudes agents can adopt in response to receiving misleading higher-order evidence? These views can also be seen as versions of two-standards views, because the distinctive attitudes they postulate come with their own rationality norms. One such account is defended by Michele Palmira (2019). Palmira agrees that agents who receive misleading higher-order evidence should reopen their deliberation. He argues that agents can then suspend judgment until they reach a new conclusion, or they can hold an attitude he calls "hypothesis" towards *p*, if they consider *p* the most promising answer to the question at hand. The solution I defend here is more general than Palmira's. Palmira doesn't specify whether any particular credences can be rationally adopted by agents upon receiving misleading higher-order evidence. Further, my account is motivated by a more general view of which attitudes are suitable for agents to adopt during ongoing reasoning processes. However, my proposal is compatible with Palmira's suggestion that agents may adopt a particular hypothesis during deliberation.

Another alternative attitude view is defended by Fleisher (2020) as a response to the self-undermining objection to the conciliatory view of higher-order evidence. Fleisher argues that conciliationists should accept an attitude towards their view he calls "endorsement", which is distinct from graded and full belief, and which "is the appropriate attitude of committed advocacy for researchers to have toward their own theories during inquiry." Fleisher argues that researchers should adopt this attitude towards their theories, and that they may hold on to it even in contexts in which they are confronted with certain kinds of challenging evidence. Being specific to research contexts, his view is less general than the view of transitional and terminal attitudes, and it not suitable for providing a response to the more general problem of misleading higher-order evidence. Also, Fleischer's view doesn't say which credences agents can adopt in light of receiving such evidence. However, my view doesn't rule out that agents could *endorse* certain claims in Fleischer's

[23]https://ralphwedgwood.typepad.com/blog/2019/07/an-unsound-argument-for-non-trivial-higher-order-evidence.html

sense. For a more detailed discussion of the differences between transitional attitudes and Palmira's notion of hypothesis as well as Fleisher's notion of endorsement, see Staffel (2019b).

Zach Barnett (2019) and Sanford Goldberg (2013) defend domain-specific views about how to respond to disagreement in philosophy. Their views address the question of which attitude philosophers can rationally take towards their views in light of widespread peer disagreement. Goldberg proposes an attitude called "speculation", which involves regarding one's view as defensible. Barnett argues for an alternative proposal, according to which one may not believe one's philosophical views, but one may set aside some of the disagreement-based higher-order evidence in one's sincere philosophical theorizing. The unmooring view differs from these views insofar as it presents a domain-general picture of how misleading higher-order evidence should impact one's deliberation and belief formation. I don't want to rule out here that evidence from disagreement should sometimes be given special treatment in particular domains of inquiry. However, proposals to this effect are not in competition with the unmooring view for giving a more general theory of the impact of higher-order evidence.

6. Conclusion

At this point it should be clear why I label my proposed view the *unmooring view* of how to respond to higher-order evidence. Receiving credible higher-order evidence that indicates that one's reasoning is likely flawed unmoores the previously settled conclusion of one's reasoning and throws the agent back into a transitional stage of their deliberation. This view preserves some of the main insights of the existing views, without incurring their costs: The first-order evidence is respected, because it determines what attitude is justified as the conclusion of the agent's reasoning. The higher-order evidence is respected, since it serves as a doxastic defeater and it influences which transitional attitudes the agent can rationally adopt before they finish their second deliberation phase. Also, the agent does not hold akratic attitudes. The unmooring view is fully compatible with standard accounts of epistemic justification, as long as we acknowledge that these accounts apply to terminal attitudes, and that the justificatory standards for transitional attitudes are different. Plausibly, this is how we were supposed to understand standard views of justification all along, so this claim merely brings out an implicit assumption of these views.

The view also generalizes nicely to related cases. What if an agent receives credible, accurate higher-order evidence that their reasoning is flawed? The unmooring view predicts that if the agent responds rationally, their attitudes will become transitional and they will reopen their deliberation just like in the misleading case. If they can find the error in their reasoning, they can now come to hold a doxastically justified terminal attitude, which they didn't have before. What about affirming higher-order evidence, i.e. evidence that indicates that the agent has reasoned correctly? In this case, the agent should keep their terminal attitude, and this attitude might even become more resilient, i.e. more resistant to possible future doxastic defeaters.

I've thus offered a view of how agents should respond to misleading higher-order evidence that avoids the pitfalls of the existing main accounts, and that integrates with standard accounts of epistemic justification and rationality. The notion of a transitional attitude that the view is built on is independently motivated by considerations of how complex deliberations proceed.

Acknowledgements:
For helpful suggestions and discussion I would like to thank Brian Talbot, Kevin Dorst, Sophie Horowitz, Robert Rupert, Caleb Perl, Heather Demarest, Matthias Steup, audiences at the Social Distance Epistemology Series, the University of Glasgow, and the University of Düsseldorf as well as an anonymous reviewer for this journal.

Bibliography

Barnett, Zach. 2019. Philosophy without Belief. *Mind* 128 (509), 109-138.

Beddor, Bob. 2015. Process Reliabilism's Trouble With Defeat. *The Philosophical Quarterly* 65 (259), 145-159.

Beddor, Bob. forthcoming. Reasons for Reliabilism. In: M. Simion and Jessica Brown (eds.), *Reasons, Justification and Defeat*. Oxford University Press.

Bogardus, Tomas. 2009. A Vindication of the Equal Weight View. *Episteme* 6 (3), 324-335.

Carey, Brandon & Matheson, Jonathan. 2013. How Skeptical is the Equal Weight View? in: D. Machuca (ed.), *Disagreement and Skepticism*, New York: Routledge, 131-149.

Christensen, David. 2007. Epistemology of Disagreement: The Good News. *The Philosophical Review* 116 (2), 187-218.

Christensen, David. 2010. Higher-Order Evidence. *Philosophy and Phenomenological Research* 81, 185-215.

Christensen, David. 2019. Formulating Independence. in: M. Skipper & A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford, Oxford University Press, 13-34.

Christensen, David & Lackey, Jennifer (eds.). 2013. *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

Coates, Allen. 2012. Rational Epistemic Akrasia. *American Philosophical Quarterly* 49 (2), 113-124.

Comesaña, Juan. 2010. Evidentialist Reliabilism. *Noûs* 44 (4), 571-600.

Comesaña, Juan. 2018. Whither Evidentialist Reliabilism? in: Kevin McCain (ed.), *Believing in Accordance with the Evidence*. Springer, 307-325.

Dallmann, Justin. 2017. When Obstinacy is a Better Cognitive Policy. *Philosophers' Imprint* 17 (24), 1-18.

DiPaolo, Joshua. 2018. Higher-Order Defeat is Object-Independent. *Pacific Philosophical Quarterly* 99, 248-269.

Dorst, Kevin. 2020. Evidence: A Guide for the Uncertain. *Philosophy and Phenomenological Research* 100 (3), 586-632.

Dunn, Jeffrey. 2015. Reliability for Degrees of Belief. *Philosophical Studies* 172 (7), 1929-1952.

Elga, Adam. 2007. Reflection and Disagreement. *Noûs* 41 (3), 478-502.

Elga, Adam. 2010. How to Disagree about how to Disagree. in: R. Feldman & T. A. Warfield, *Disagreement*, Oxford: Oxford University Press, 175-186.

Feldman, Richard. 2009. Evidentialism, Higher-Order Evidence, and Disagreement. *Episteme* 6 (3), 294-312.

Feldman, Richard & Warfield, Ted A. (eds.). 2010. *Disagreement*. Oxford: Oxford University Press.

Firth, Roderick. 1978. Are Epistemic Concepts Reducible to Ethical Concepts? in: A. I. Goldman & J. Kim (eds), *Values and Morals*, Dordrecht: D. Reidel, 215-229.

Fleisher, Will. 2020. How to Endorse Conciliationism. *Synthese*, online first.

Goldberg, Sanford. 2013. Defending Philosophy in the Face of Systematic Disagreement. in: D. E. Machuca (ed.), *Disagreement and Skepticism*, Routledge, 277-294.

Goldman, Alvin. 1979. What is Justified Belief? in: G. Pappas (ed.), *Justification and Knowledge*, Dordrecht: D. Reidel, 1-23.

Good, I.J. 1967. On the Principal of Total Evidence. *The British Journal for the Philosophy of Science* 17 (4), 319-321.

Hazlett, Alan. 2012. Higher-Order Epistemic Attitudes and Intellectual Humility. *Episteme* 9 (3), 205-223.

Horowitz, Sophie. 2014. Epistemic Akrasia. *Noûs* 48 (4), 718-744.

Icard, Thomas. 2018. Bayes, Bounds, and Rational Analysis. *Philosophy of Science* 85, 79-101.

Isaacs, Yoaav. forthcoming. The Fallacy of Calibrationism. *Philosophy and Phenomenological Research.*

Kelly, Thomas. 2005. The Epistemic Significance of Disagreement. *Oxford Studies in Epistemology* 1, 167-196.

Kelp, Christoph. 2019. How To Be A Reliabilist. *Philosophy and Phenomenological Research* 98 (2), 346-374.

Lasonen-Aarnio, Maria. 2020. Enkrasia or Evidentialism? Learning to Love Mismatch. *Philosophical Studies* 177, 597-632.

Lord, Errol. 2020. Suspension of Judgment, Rationality's Competition, and the Reach of the Epistemic. in: Sebastian Schmidt and Gerhard Ernst (eds.), *The Ethics of Belief and Beyond. Understanding Mental Normativity.* Routledge, 126-146.

Lyons, Jack C. 2016. Goldman on Evidence and Reliability. In: H. Kornblith and B. McLaughlin (eds.), *Goldman and his Critics*. Blackwell, 149-177.

Neta, Ram. 2019. The Basing Relation. *The Philosophical Review* 128 (2), 179-217.

Palmira, Michele. 2019. How to Solve the Puzzle of Peer Disagreement. *American Philosophical Quarterly* 56 (1), 83-96.

Pettigrew, Richard. 2018. What is Justified Credence? *Episteme*, online first.

Pryor, James. 2018. The Merits of Incoherence. *Analytic Philosophy* 59 (1), 112-141.

Russell, Stewart & Wefald, Eric. 1991. Principles of Metareasoning. *Artificial Intelligence* 49, 361-395.

Silva, Paul Jr. 2017. How Doxastic Justification Helps Us Solve the Puzzle of Misleading Higher-Order Evidence. *Pacific Philosophical Quarterly* 98 (S1), 308-328.

Skipper, Mattias & Steglich-Petersen, Asbjorn. 2019. *Higher-Order Evidence: New Essays*. Oxford: Oxford University Press.

Sliwa, Paulina & Horowitz, Sophie. 2015. Respecting All the Evidence. *Philosophical Studies* 172 (11), 2835-2858.

Smithies, Declan. 2012. Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research* 85 (2), 273-300.

Smithies, Declan. 2019. *The Epistemic Role of Consciousness*. Oxford: Oxford University Press.

Staffel, Julia 2019a. How Do Beliefs Simplify Reasoning? *Noûs* 53 (4), 937-962.

Staffel, Julia. 2019b. Credences and Suspended Judgments as Transitional Attitudes. *Philosophical Issues* 29 (1), 281-294.

Staffel, Julia. 2020. Pro Tem Rationality. forthcoming in *Philosophical Perspectives* 35.

Tang, Weng Hong. 2016. Reliability Theories of Justified Credence. *Mind* 125 (497), 63-94.

Titelbaum, Michael. 2015. Rationality's Fixed Point (Or: In Defense of Right Reason). *Oxford Studies in Epistemology* 5, 253-294.

Titelbaum, Michael. 2019. Return to Reason. in: M. Skipper & A. Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, Oxford, Oxford University Press, 226-245.

van Inwagen, Peter. 1996. It is Wrong, Always, Everywhere, and for Anyone, to Believe Anything, Upon Insufficient Evidence. in: J. Jordan & D. Howard-Snyder (eds.), *Faith, Freedom, and Rationality*, Hanham, MD: Rowman and Littlefield, 137-154.

Van Wietmarschen, Han. 2013. Peer Disagreement, Evidence, and Well-Groundedness. *The Philosophical Review* 122 (3), 395-425.

Vul, Edward, Goodman, Noah, Griffiths, Thomas L. & Tenenbaum, Joshua B. 2014. One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science* 38, 599-637.

Weisberg, Jonathan. 2020. Could've Thought Otherwise. *Philosophers' Imprint* 20 (12), 1-24.

Williamson, Timothy. 2011. Improbable Knowing. in: T. Dougherty (ed.), *Evidentialism and Its Discontents,* Oxford: Oxford University Press, 147-64.