

Are there transitional beliefs? – I think so?

Julia Staffel

Draft for the Rutgers Epistemology Conference (April 10, 2024)

Please don't cite without permission. Comments welcome!

Introduction

A question that has gathered much interest in epistemology recently is whether it can ever be rational to keep inquiring into a question once one has adopted a belief that answers it. For example, if Detective Fletcher believes that Manny committed the murder, is it rationally permissible for her to keep inquiring into who committed the murder? Friedman (2019) has prominently argued for a negative answer to this question. She proposes the norm DBI (Don't believe and inquire): One ought not inquire into/have an interrogative attitude towards a question Q at some time t and believe a complete answer to Q at t . Hence, in order to keep inquiring, one needs to stop believing an answer to Q and adopt an interrogative attitude instead, or so Friedman claims. Her interlocutors have offered various counterarguments to DBI, interrogating whether it is really a true norm of inquiry.¹

I am interested in a related, but slightly different question here, which has not gathered any attention, but will help us better understand the nature of belief and its relation to inquiry and deliberation. The question I will be thinking about is: Is it ever rationally permissible to believe something prior to concluding one's deliberation by settling on an answer to Q ? My question differs from the more commonly discussed one, insofar as the former asks about the rationality of believing that p *before* settling on p as the answer to Q , while the latter asks about the rationality of continuing to inquire into Q *after* coming to believe that p is the answer to Q .

My question might initially strike some readers as very strange – if we adopt the common lore that to believe p is to settle the relevant question Q to which p is a complete answer, then it might not even seem possible to believe p without thereby concluding that p is the answer to Q . Or, at least, this might only seem possible for an agent with a fragmented state of mind, who somehow doesn't "bring together" their belief in p and their inquiry into Q .

The question makes more sense once we take into consideration a distinction between doxastic attitudes that are formed while an agent is in the process of deliberating about a question, which I call *transitional attitudes*, and doxastic attitudes that are formed as conclusions of such deliberations, which I call *terminal attitudes*. As I have argued in earlier work, epistemological theorizing greatly benefits from making this distinction, because doxastic attitudes play different roles in our reasoning depending on when during our deliberations they are formed.

¹ Some have argued that it can be rationally permissible to come to believe that p is the answer to Q , and to keep inquiring into Q while one holds that belief. For instance, Falbo (2023) argues that the goal of inquiry is epistemic improvement, and that one can sometimes rationally seek epistemic improvement even if one already believes an answer to Q . Relatedly, Willard-Kyle (2022) argues against a range of norms that claim that agents should always stop inquiring once they have reached a particular epistemic benchmark, like being in a position to know. Others have argued that DBI can't be both correct and normative: Lee (2023) argues that if we interpret DBI as applying to occurrent beliefs, it's impossible to violate, and hence, while descriptively true, it doesn't state a norm.

With this distinction in hand, we can better articulate the two questions from above: The question that has received ample discussion so far can be rephrased as: Is it ever rationally permissible to keep inquiring into the answer to Q once one has adopted a *terminal* belief that fully answers Q? By contrast, the question I am interested in is: Is it ever rationally permissible to *transitionally* believe something before one has concluded one's deliberation, i.e., before one has settled on an answer to Q?

Since epistemologists have previously identified beliefs with what I call *terminal beliefs*, the question of whether there could be such a thing as transitional beliefs couldn't be formulated within their taxonomy of doxastic attitudes. But now that it can be asked, trying to answer it can shed new light on questions about the nature of belief, for example about whether they are really as stable and settled as is usually assumed. We can ask: Is it because these attitudes are beliefs, or is it because they are terminal attitudes, that they have the distinctive features often attributed to them? Could we have transitional beliefs that lack some of these features but still be considered beliefs? I will argue that it is possible for rational agents to hold transitional beliefs. Further, I will show that many common claims about what beliefs are don't identify important features of belief itself, but of terminal attitudes more generally.

In what follows, I will first introduce and motivate the distinction between terminal and transitional attitudes. I will then present some examples that provide preliminary evidence for the existence of rational transitional beliefs, and offer various more and less conservative ways of accounting for the data I present. In the subsequent sections, I will consider and attempt to dispel two challenges to the claim that we can have transitional beliefs: a descriptive challenge based on common views about the nature of belief, and a normative challenge based on the alleged impermissibility of having akratic doxastic attitudes.

1. Attitudes in Deliberation: Transitional and Terminal

In order to motivate the distinction between transitional and terminal attitudes, it is helpful to first differentiate between inquiry on the one hand, and reasoning or deliberation on the other hand.² Inquiry is usually understood as an activity that we undertake in order to answer a question we're interested in, and it can involve a number of different components, such as seeking out and learning new information, thinking about one's evidence, formulating new hypotheses, revising existing attitudes, and so on. Reasoning, or deliberation, by contrast, is the activity of thinking about the evidence one has, and how it bears on the question at hand. It can be a component of inquiry, but it doesn't involve, for example, gathering and learning new information. Not every instance of answering a question involves an extended period of deliberation. For example, when I figure out the answer to a simple math problem, or when I infer that my classes will be poorly attended today because the weather is bad, these inferences are carried out by my mind quickly and automatically. By contrast, there are also many cases in which it is not obvious how my information bears on a question, and the way to figure it out is by reasoning. Examples of this are commonplace both in academic and everyday contexts. In my philosophical work, I might spend some time thinking

² For more detailed discussions of my view on terminal and transitional attitudes, see Staffel (2019, 2021, 2023, 2024a).

about how to adjust my view in order to deal with a tricky counterexample. In everyday life, it might take me some time to decide on a phone or healthcare plan, because I need to think about how the subtle differences between the plans will play out in my particular circumstances. Reasoning or deliberation, so understood, is what I will be concerned with in what follows.³

With these clarifications in hand, I will now introduce a couple of examples of complex deliberation that will help introduce the distinction between transitional and terminal attitudes:

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 5%.

Phone Plan:

Bill has just moved to the US and needs to sign up for a cellphone plan. He has carefully studied the various offerings, but the plans are hard to compare because their pricing structures are so different. Initially, he has low credence that one of the more expensive plans is the best option for him, because he doesn't use his phone very often. But, upon further reflection, it occurs to him that cheaper plans nickel and dime customers for using the phone while traveling. This could get expensive for him very quickly, so he ends up concluding that one of the higher-priced plans with some included international credits is going to work best for him.

In both cases, an agent is engaged in a reasoning process aimed at answering a particular question based on the relevant information they possess (Who committed the murder? Which is the best phone plan for me?). Each agent ultimately arrives at a conclusion that is, let's stipulate, the rational attitude for them to arrive at. But notice that each agent also adopts a variety of credences in different candidate answers to the question while they are deliberating that are different from the attitudes that constitute the conclusions of their reasoning. Those attitudes also seem rational for the agents to have, given the stage of deliberation they're at: before Fletcher sees through the framing attempt, her evidence makes it appear quite likely that Fred did it. In light of this, it is reasonable for her to have at least a preliminary high confidence at this stage of her deliberation that Fred is the murderer. It would be much less fitting for her to have, for instance, a low preliminary degree of confidence that Fred did it, if she really has no inkling yet that he's been set up by Manny. In Bill's case, it makes sense for him to initially have low confidence that he needs an expensive phone plan, because he's only considered his domestic phone usage at that time, and

³ For ease of exposition, I will assume that the information that agents reason about stays fixed, and that they don't learn any additional empirical evidence before they conclude their deliberation. This is of course not always true in real life, but nothing hangs on this simplifying assumption. For an argument against the view that reasoning itself changes what the agent's evidence supports, see Staffel (2024b).

he hasn't factored in that he will occasionally need to use his phone while he travels. Before he considers his international phone use, Bill has very little reason to think that he needs an expensive cellphone plan.

In short, these two cases illustrate a common phenomenon we observe in cases of complex deliberation, which can be described with these two claims:

(I) The attitudes that are rational for the agent to adopt before the reasoning is completed can differ from the attitudes that are rational when the agent has finished deliberating.

(II) While in the process of reasoning, some attitudes the agent could have are more rational than others they could have.

Unfortunately, as I have discussed in detail elsewhere, (I) and (II) cannot be accommodated by any of the standard theories of epistemic rationality or justification (Staffel 2021, 2024a,b). Those theories are geared towards telling us whether the concluding attitudes generated by reasoning processes are rational or justified. These concluding attitudes have so far been the sole focus of normative epistemological theorizing about doxastic attitudes, such as beliefs, credences, and suspensions.

Consider evidentialism, which says, roughly, that a doxastic attitude is justified or rational just in case it is supported by the agent's total relevant evidence. Evidentialism can explain why the agents in our cases arrive at justified conclusions, i.e., very low confidence that Fred is the killer, and a belief that a more expensive phone plan is best for Bill, since those attitudes are justified by the agent's total evidence. But this means that any attitude that is meaningfully different from those will thereby count as unjustified or irrational according to evidentialism, since it is not supported by the agent's total evidence.⁴ Hence, evidentialism can't account for the judgment that the attitudes that Fletcher and Bill have during early stages of deliberation are rational, and that those attitudes are more rational for them to have at that point than the attitudes they will ultimately settle on, without yet seeing the rationale for adopting them.

The same point applies, *mutatis mutandis*, to other theories of justification. For instance, reliabilism says, roughly, that a justified or rational doxastic attitude is one that is the output of a sufficiently reliable reasoning process. But the attitudes characterized by (I) and (II) are not the outputs of sufficiently reliable reasoning processes, hence, reliabilism can't account for their rationality. For another example, Bayesianism says, roughly, that a credence is justified or rational just in case it was formed by conditionalizing on one's total evidence, starting with rational priors. Again, this delivers the result that our agents should have the credences that they end up with at

⁴ This point holds independently of whether we endorse a permissive version of evidentialism that allows for a range of rational responses to the evidence, or a stricter version that only allows one unique rational response to the evidence.

the end of their deliberations, but we have no way of accounting for the rationality of their earlier credences.⁵

It seems then, that in order to account for (I) and (II), we need to formulate distinct, weaker standards of rationality that don't have some kind of total evidence or completion requirement built into them. But one might worry that this is an ad hoc move. If the doxastic attitudes we form while we're deliberating are not in any way different from the ones we adopt as conclusions of deliberation, then applying two different rationality standards to them seems arbitrary. After all, we don't think rationality standards vary by the time of day or the day of the week, so why should they differ depending on the stage of deliberation?

To see why introducing distinct rationality standards is not so arbitrary after all, it is useful to observe more closely how the roles our doxastic attitudes can play differ depending on where they appear in a reasoning process. I will use the following terminology: a doxastic attitude is a *terminal attitude* just in case it is adopted by an agent when they have finished deliberating about the question they want to answer. A terminal attitude is what we also commonly call a conclusion of reasoning, but it need not be a belief, it could also be a credence, or a neutral attitude, depending on what the agent thinks is the warranted attitude to arrive at given the information they possess. Notice that properly finishing a deliberation is not the same as stopping it. Sometimes we abandon a reasoning process before reaching a conclusion. Merely stopping a reasoning process is not sufficient for giving the doxastic attitudes the agent has at that point the status of being terminal attitudes, rather, the agent must deem the reasoning process *properly completed*. In our examples, Fletcher's 5% credence that Fred is the murderer and Bill's belief that a more expensive phone plan is best for him are terminal attitudes.

By contrast, *transitional attitudes* are attitudes towards potential answers to the question under consideration that the agent forms at the beginning and intermediate stages of a reasoning process. They reflect how plausible different candidate answers look in light of the reasoning that the agent has completed up to that point. However, not all attitudes an agent forms while deliberating are thereby transitional attitudes. For example, Bill might form the belief at an intermediate stage of his deliberation that phone plan A is cheaper for domestic calls than phone plan B. This belief reflects his considered opinion, and counts as a terminal attitude, because at that point, Bill is no longer deliberating which of the two plans is cheaper for domestic calls. This belief is an important premise in his larger deliberation about which plan is best overall. An attitude is transitional just in case it reflects the agent's *non-final* take on the matter at hand.

Whether an attitude is transitional or terminal is entirely dependent on a particular agent's psychological state. The agent (or their cognitive system, this need not be conscious) decides whether a reasoning process has delivered a sufficiently justified answer to the question at hand. If the agent thinks that it has, the resulting attitude is terminal, if not, transitional. This view is informed by empirical models of complex deliberation and metacognition, which suggest the

⁵ I don't expect my readers to be convinced by these quickly sketched arguments. They are just intended to outline the problem posed by (I) and (II) for standard theories of rationality and justification. For a much more detailed discussion, I invite readers to consult Staffel (2021, 2024a,b).

following view of how an agent manages extended reasoning processes: The reasoner carries out some strategy to solve a given problem, while simultaneously employing metacognitive monitoring and control mechanisms that help determine whether the problem-solving strategy is being carried out correctly or needs to be adjusted, and whether it has delivered an acceptable conclusion. As the agent deliberates about how their truth-relevant evidence bears on answering the question under consideration, their metacognitive monitoring system assesses whether the target level of confidence in the quality of their reasoning has been reached. Once the reasoning seems sufficiently good, the reasoning process is completed and the resulting answer is accepted as a terminal attitude. However, if the agent can't establish a full or partial answer that meets the desired level of reasoning quality (which may also be adjusted during the reasoning process), they'll give up and fail to reach a terminal attitude (see, e.g., Ackerman 2014, Thompson and Johnson 2014, Ackerman & Thompson 2017).

While terminal attitudes are end points of reasoning processes, they need not be end points of inquiry more broadly construed. They are still subject to being changed and updated if the agent learns new relevant information. They can also revert to having the status of transitional attitudes if the agent decides to reopen or double-check the reasoning process they had previously deemed finished.

Having now introduced this distinction, we can observe some interesting descriptive differences between transitional and terminal attitudes. Terminal attitudes are typically available for unrestricted further use as premises in reasoning and decision making, and as bases for actions and assertions. This is widely accepted in the literature (see section 3 for further discussion and references). This is not true, however, of transitional attitudes. While they play a role in reflecting the current status of the agent's deliberation about a specific question, and help guide and structure that reasoning process, they are typically not available as premises for reasoning and decision making outside of that immediate reasoning context. For example, Bill wouldn't decide to increase his monthly retirement savings based on his transitional high credence early in his deliberation that a cheaper phone plan is best for his needs. (By contrast, he might very well decide to decrease his monthly savings based on his *terminal* belief that a more expensive phone plan is best for him.) Further, agents don't typically assert transitional attitudes either, or if they do, they tend to indicate their transitional status in some way. For example, if Fletcher were asked early in her deliberation whether Fred is the killer, she wouldn't just say: "Yes, he almost certainly is." Rather, she would indicate that this is not her final take on the matter, by saying something like "It looks likely to me right now that he is, but I haven't finished thinking about it", or perhaps just "I think he probably is?" By adding some type of hedge, through her tone or word choice, the agent indicates that the listener should not interpret the utterance as expressing the agent's considered take on what their evidence supports.

It makes sense that agents typically don't rely on transitional attitudes in the same manner in which they rely on terminal attitudes – after all the transitional attitudes don't live up to the agent's own standards of justification that they consider sufficient for terminating the deliberation. However, this is not to say that agents never rely on them. Sometimes, we're not in a position to finalize our reasoning to our satisfaction, for example because we're under time pressure or because

we don't know how to properly solve a hard problem. In those cases, we might still be forced to make a decision or act, and our transitional attitudes are our best (if imperfect) take on the matter at that point. For example, on a timed exam, you might not be able to verify your answers to your satisfaction, so it's best to use your transitional attitudes to answer the question (compared to simply guessing or not answering at all). How inclined an agent is to rely on their transitional attitudes will also depend on how far along in their deliberation they are. If an agent deems their reasoning to be almost completed, then their transitional attitudes are close to being justified enough to be "promoted" to being terminal, and those attitudes will be more available for being relied upon than transitional attitudes that are formed at the beginning and in early stages of deliberation.⁶

The differences I have just pointed out between terminal and transitional attitudes are descriptive differences that concern the roles typically played by these attitudes in reasoning. Once we recognize that they differ in their roles, we have a proper basis for the claim that they differ normatively as well. The appropriate normative criteria for evaluating the rationality of attitudes depend on what roles those attitudes play – they tell us what these attitudes need to be like in order to play their roles *well*. For example, the norms for rational intentions and rational beliefs differ, because beliefs and intentions play different roles. Similarly, given that terminal attitudes are used as bases for actions, assertions, and further reasoning, while transitional attitudes play more restricted roles in structuring an ongoing deliberation and reflecting its preliminary results, it makes sense that the former are subject to more stringent rationality requirements than the latter. The standard views of epistemic rationality or justification, which include some kind of reasoning completion or total evidence requirement, are well suited for terminal attitudes, but as I pointed out above, they are too strong for evaluating transitional attitudes. For transitional attitudes, we need to formulate weaker rationality conditions. As a first approximation, transitional attitudes are rational at a particular stage of reasoning just in case they reflect the evidence the agent has considered up to that point, and the agent's take on how that evidence bears on the question at hand. I have developed this idea in much more detail in Staffel (2021, 2024a,b) but this rough sketch of the view is all we need in what follows.

2. Transitional Beliefs? Cases and Possible Ways of Accounting for Them

In categorizing doxastic attitudes, epistemologists tend to either adopt a graded or a categorical taxonomy: according to the former, our beliefs come in degrees. They can range all the way from certainty that something is true to certainty that something is false. A common way of modeling degrees of belief is via numerical credences in the zero-one interval, although those numbers are of course just part of the representation of our attitudes, they don't actually exist in our minds. According to the categorical taxonomy, an agent either believes something, disbelieves it, or has an agnostic attitude about it.

The question I am interested in is: Which of these attitudes can serve as transitional attitudes? It is standardly assumed that all these attitudes can serve as *terminal* attitudes: a process of reasoning can conclude in a credence of any strength, an agnostic attitude, a belief, or a disbelief,

⁶ For more discussion, see Staffel (2024, ch. 5).

depending on what the agent thinks is warranted by their deliberation about what their evidence supports. For transitional attitudes, non-extreme credences and an agnostic stance seem like the least controversial options.⁷ After all, transitional attitudes are adopted before the agent settles on any particular terminal attitudes, so it seems natural that the agent's best take on how to answer the question under consideration at that point should leave multiple options open. Hence, on this line of thinking, the agent may assign an agnostic attitude to the different candidate answers or assign them non-extreme degrees of belief depending on how plausible they seem at that point.⁸

However plausible one might think it is that we can have terminal and transitional credences and agnostic attitudes (and there is certainly much more to be said about this, see Staffel 2019, 2023), it is arguably much less plausible that there can be transitional and terminal versions of beliefs and disbeliefs.⁹ Belief is typically seen as a kind of settling attitude, which commits the believer to accepting something as true, and to act and talk on this basis. This picture fits well with the claim that beliefs can be terminal attitudes, but it is at odds with the idea that we can have transitional beliefs. The whole point of having a transitional attitude is that it is *not* settled and that we typically *don't* use it as a basis for assertion and action. So how could there be transitional beliefs?

Let's first think about this from a theoretical perspective. What kinds of reasoning situations could in principle give rise to transitional beliefs? I said earlier that the function of transitional attitudes is to reflect how plausible different candidate answers look at a given stage of deliberation, and to help guide and structure further deliberative moves. Further, the empirically informed view of complex deliberation I laid out above suggests that there are two processes involved – one process that operates at the first-order level, in which the agent evaluates how their truth-relevant or first-order evidence bears on the question at hand, and a higher-order monitoring and control process that checks and steers the execution of the first-order process. Transitional attitudes reflect how things look with respect to Q given the agent's current insight into their evidence, so they are best construed as reflecting the preliminary results of the agent's first-order, or truth-relevant reasoning.

Given this picture, we can then distinguish two kinds of results that the agent's first-order reasoning might deliver at a particular stage of reasoning: a categorical result, or a non-categorical result: Suppose there is a partition of complete answers to Q. A (partially or fully completed) process

⁷ I don't make any assumptions here about the relationships between credences and categorical attitudes. I think they can both be correctly ascribed to agents at the same time. On the most plausible views, there is no straightforward reductive relationship between them, but my arguments in this paper don't depend on any such claim (see, e.g. Friedman 2013, Weisberg 2020).

⁸ Notice that I am not assuming here that transitional credences and transitional suspensions are somehow fundamentally different kinds of attitudes from their terminal counterparts. While they play somewhat different roles, they are crucially similar in key regards. For example, both transitional and terminal credences encode the agent's uncertainty about some matter, they are sensitive to evidence and can change in response to learning and deliberation, and so on. Further, when the agent decides that they are ready to conclude their reasoning, they don't thereby abandon their transitional credences and replace them with terminal credences. On a more plausible and parsimonious picture, the agent's existing transitional credences undergo a change in status due to the agent deeming them sufficiently well supported to promote them to being terminal attitudes. We can say similar things about transitional and terminal versions of agnostic attitudes.

⁹ I'll focus on beliefs in what follows, but all the points also apply to disbeliefs, insofar as one conceives of them as a separate attitude from believing that something is not the case.

of reasoning delivers a categorical answer if it singles out one element of the partition as the answer to Q . It delivers a non-categorical answer if it doesn't single out one element as the correct answer. If multiple answers remain in play at a given stage of deliberation, the agent's reasoning might provide different degrees of support for different answers.

In the examples we previously considered, the agents' preliminary deliberations didn't deliver categorical answers. Rather they merely made some answers look more likely than others. But we can easily imagine cases in which an agent's preliminary deliberations deliver a categorical answer, even if the agent is not ready yet to settle on this answer as a terminal attitude because they don't deem their reasoning sufficiently good quite yet. These are precisely the types of cases that are candidates for permitting the adoption of transitional beliefs. Here are a couple of examples that demonstrate this possibility:

Dinner bill:

David and his five friends are out to dinner. They've agreed to split the bill evenly between them and to tip 20%. At the end of the meal, the bill comes and David sees that the bill with tax is \$182.50. After a quick calculation in his head, it seems to him that everyone owes \$36.50. But, since he's in charge of making sure the bill gets paid correctly, he double checks his math more carefully to make sure he gets it right. He confirms his earlier result and collects \$36.50 from each dinner guest to pay the bill.

Logic problem:

Tilda is a beginning logic student who is learning truth tables for propositional logic. She has mastered the standard truth table method for determining whether a formula is a tautology. In her last class, she learned a new method: assume the whole formula is false, and see if you can fill in all the relevant truth values without a contradiction. If you can do so, the formula is not a tautology. Tilda starts on a formula and tries the new method. Her attempt to fill in all the truth values seems to work out consistently, suggesting that the formula is not a tautology. Since she doesn't yet feel confident about her mastery of this method, she writes out the full truth table and discovers that she made a mistake. The formula is a tautology after all.

These two cases differ from the original ones insofar as the agents' preliminary first-order reasoning delivers a categorical answer to the question at hand: In David's case, it's that each person owes \$36.50, and in Tilda's case, it's that the formula is not a tautology. If there is such a thing as a rational transitional belief, this is the kind of case in which an agent could have one.

But of course, whether we can attribute rational transitional beliefs to Tilda and David is precisely what is under discussion, so we also need to spell out what the alternatives to this view might be. In particular, it seems to me that the following reactions might be common among my readers: One possible take on what's going on in these examples is that the agents have some sort of binary attitude before their reasoning is completed, but that attitude is not a belief. Drawing from the standard inventory of binary non-belief attitudes, one might try to categorize them, for example, as acceptances, hypotheses, or suppositions. Another possible reaction might be that the

agents should be agnostic, or have non-extreme credences in the claims delivered by their reasoning. For example, one might think that the agents can rationally have a transitional high credence that the bill is \$36.50, and that the formula isn't a tautology, but not a transitional belief. This reaction is likely driven by the observation that the agents have significant *higher-order* uncertainty about the correctness of these claims, given their evaluation of the quality of their reasoning up to that point. Incorporating this higher-order uncertainty into their first-order attitudes would lead the agents to have transitional agnostic attitudes or non-extreme credences.

We thus need to adjudicate between three possible views of how to categorize the attitudes our agents have in situations in which their incomplete first-order reasoning delivers a categorical answer to the question under consideration (of course there could be many others, but I think most epistemologists would gravitate to one of these):

Option 1: It is possible for agents to have rational transitional beliefs, and those are the attitudes we should ascribe to David and Tilda.¹⁰

Option 2: It is not possible for agents to have rational transitional beliefs, and the attitudes agents can rationally have in those cases are a different kind of categorical attitude, such as acceptance, hypothesis, or supposition.

Option 3: It is not possible for agents to have rational transitional beliefs, and the attitudes agents can rationally have in cases like David's and Tilda's are credences and agnostic attitudes.

In what follows, my aim is to make the best possible case for option 1. In doing so, I will discuss two main challenges to endorsing it. One of these is a descriptive challenge, which claims that reflection on the nature of belief shows that the attitudes that can suitably be ascribed to David and Tilda can't possibly be beliefs, because they don't share enough features with beliefs, properly understood. The second challenge is a normative challenge, which claims that having transitional beliefs would lead to epistemic akrasia, which is normatively objectionable. On this view, even if it's possible to have transitional beliefs, those beliefs would be irrational. In what follows, I will develop these challenges in turn and discuss how a proponent of option 1 can respond to them.

3. The Challenge from the Nature of Belief

In the literature on the nature of belief, philosophers have pursued a variety of strategies to characterize belief and to delineate it from other doxastic attitudes. They have, for instance, analyzed the functional role of belief in our cognitive lives, the role of belief in inquiry, the phenomenal character of belief, and the distinctive aim of belief. I can't adjudicate between the merits of these different approaches here, nor can I attend to the subtle variations within versions

¹⁰ Of course, someone could have the view that it is possible to have rational transitional beliefs, but that I have not given any cases in which this occurs. I am not sure what kinds of better cases one might have in mind that are distinct from the ones I am proposing, so I will not discuss this option further.

of a single approach. Instead, I will try to characterize each of them in broad brushstrokes and ask if it leaves room for the existence of transitional beliefs.

3.1 Belief as a Bundle of Characteristic Dispositions

On a very popular view of belief, to believe that p is to be in a mental state that has a number of characteristic phenomenal and functional features (see, e.g. Alston 1996, Chalmers 1996, Schwitzgebel 2002). We get a list of such features from Alston (adapted from Alston 1996, p. 4):

Affirmation: If S believes that p , then if someone asks S whether p , S will have a tendency to respond in the affirmative.

Feeling: If S believes that p , then if S considers whether p , it will feel to S that p is true.

Inference: If S believes that p , then S will tend to believe propositions that S takes to follow from p .

Premising: If S believes that p , then S will tend to use p as a premise in theoretical and practical reasoning.

Surprise: If S believes that p , then if S learns that $\sim p$, S will tend to be surprised.

Action: If S believes that p , then S will act in ways that would be appropriate if p were true, given S 's desires and other beliefs.

Involuntariness: S 's belief that p is not under S 's voluntary control. S cannot choose whether to believe, disbelieve, or be agnostic about p .

This incomplete list tells us the characteristic tendencies or dispositions that are associated with belief. I won't discuss here whether any of these are necessary for belief, or more or less central. Rather, I will ask what this list can tell us about the prospects for the existence of transitional beliefs.

Suppose option 1 from above is correct: If an agent has a transitional belief that p , this means that the agent's unfinished first-order deliberation has returned a categorical answer to the question at hand, and thus a transitional belief in this answer is the attitude that best reflects the results of the agent's reasoning up to that point. The agent does not hold a terminal belief in the answer, because they don't trust their reasoning up to that point sufficiently to conclude their reasoning. Do transitional beliefs, so described, have the features we should expect them to have, given Alston's list?

The following characteristic features from Alston's list apply to it pretty straightforwardly: Feeling, Surprise, Involuntariness, and Inference. Consider Feeling first.¹¹ When we reason about an issue and our deliberation up to that moment straightforwardly points to a particular answer, that answer will seem true to us in light of this reasoning. We might refrain from immediately concluding our reasoning with this answer because we have second-order qualms about whether our reasoning has been sufficiently trustworthy. But this doesn't change the fact that the answer

¹¹ Some accounts of belief privilege the feeling of truth or conviction as the most central aspect of belief. See, e.g., Cohen (1992) and Smithies (2023).

our reasoning suggests at that time feels “truthy” to us, and alternative answers don’t. Once a transitional attitude becomes a terminal attitude, the attitude receives the stamp of approval of having been arrived at by sufficiently good reasoning, but this does not necessarily change the cognitive feeling of truth associated with the attitude.¹²

Next, consider Surprise: while it’s certainly not *as* surprising to learn that the result of one’s preliminary reasoning has been mistaken as it is surprising that one has reached a false conclusion, an element of surprise is surely present. Consider the example of Tilda from the previous section, who discovers by doing a full truth table that her initial attempt to use the shortcut method delivered the wrong result. While she is aware that she has not fully mastered the shortcut method yet (that’s why the double-checks!), it would be very natural for her to be surprised to discover the mistake.

The third feature that applies to transitional beliefs is Involuntariness. It is commonly assumed in the literature that terminal beliefs cannot be adopted at will, or at least, that the degree to which we have voluntary control over these attitudes is much smaller than for other attitudes such as acceptance or supposition (though see Steup 2017 for discussion). Transitional beliefs (and other transitional attitudes) are similar to terminal beliefs in that they are responsive to evidence and deliberation, and changes in them tend to result from the agent’s consideration of truth-relevant reasons. There’s of course much to be said about exactly how this works for terminal beliefs, the important point for us is that since transitional and terminal beliefs are ultimately generated by the same reasoning mechanisms, the way they can be changed and controlled is not interestingly different.

The fourth feature that applies to transitional beliefs is Inference, albeit in a somewhat limited manner. If at any point during the agent’s deliberation, they form a transitional attitude of some sort towards p , they will be inclined to also adopt other transitional attitudes that follow from these. This makes sense, because those attitude would also be supported by the agent’s deliberation up to that point in time. For example, in our two cases above, David will transitionally believe that the bill is not \$37 per person, etc., and Tilda will transitionally disbelieve that the formula is a tautology. However, agents would likely not be inclined to put much effort into drawing out the consequences of their transitional beliefs (unless doing so would seem useful in their further reasoning about the question under consideration), since they are aware that these attitudes are merely preliminary, which means that spending much effort on figuring out their entailments seems like a waste of time.

While these four features apply to both transitional and terminal beliefs, the remaining three features from Alston’s list don’t straightforwardly apply to transitional beliefs: Affirmation, Premising, and Action. This is not surprising at all. When I first introduced the distinction between

¹² Notice also that the Feeling feature applies, *mutatis mutandis*, to other doxastic attitudes: whatever the feeling of uncertainty is that accompanies having a terminal agnostic attitude or a terminal non-extreme degree of confidence similarly accompanies the transitional versions of those attitudes. This further speaks in favor of the claim that transitional and terminal versions of an attitude are adequately characterized as being attitudes of the same general type.

transitional and terminal attitudes, one of the main differences between them I pointed out was that transitional attitudes are typically not available as bases for assertion, action, and as premises in further reasoning. This is entirely understandable, since they don't constitute the agent's considered opinion, so the agent is not inclined to rely on them unless forced to do so. This is the case for all transitional attitudes, including credences and agnostic attitudes.

Since the category of a transitional attitude was not previously part of the standard taxonomy of attitudes in epistemology, any theorizing about the nature of attitudes has automatically focused on the terminal versions of attitudes. Alston's list was drawn up to characterize *terminal* beliefs, so it's not surprising that some of the features on the list can be best attributed to the fact that the attitudes under consideration are *terminal*, rather than to the fact that they are beliefs. If this is right – that Affirmation, Premising and Action are better thought of as features of attitudes that are terminal, rather than as features of any type of belief, then the fact that they don't apply to transitional beliefs is not a strike against option 1 from above.

I will offer two further arguments for the claim that Affirmation, Premising and Action are best thought of as applying to terminal attitudes, rather than beliefs. The first argument relies on a continuity with other attitudes that come in transitional and terminal versions. As I pointed out earlier, doxastic attitudes that somehow encode uncertainty, such as agnostic attitudes and non-extreme credences, plausibly come in both transitional and terminal versions. Yet, only the terminal versions of these attitudes are typically used unrestrictedly as inputs for further theoretical and practical reasoning, and as bases for assertion and decision making. A case for the rationality of this practice can be made, for instance, by appealing to Good's theorem of the value of information (Good 1967). Good's theorem says that it is beneficial for the expected value of one's decisions to incorporate free information in one's decision making. Hence, given that only terminal credences have, by the agent's lights, fully absorbed the information available to them, agents should generally rely on terminal credences in decision making, rather than transitional credences that may not fully and properly reflect their total evidence.

Given that we see the same pattern of applicability of Affirmation, Premising, and Action with respect to transitional and terminal credences, agnostic attitudes, as well as beliefs, this is good evidence that those features don't apply to terminal beliefs because they are beliefs, but because they are *terminal* attitudes. If they were key features of beliefs specifically, it would be very surprising to see that they also apply to terminal credences and agnostic attitudes, but fail to apply to the transitional versions of these attitudes. Hence, the fact that these features don't apply to transitional beliefs is not a good reason to disqualify these attitudes from counting as beliefs. Rather, this is to be expected if what these features really capture is the terminalness of an attitude.

To strengthen the idea that Affirmation, Premising and Action apply to attitudes in virtue of their status as terminal attitudes, we can also consider cases in which the agent's attitude remains the same throughout their deliberation. For example, suppose one of David's friends is late for dinner and has asked David to order an appetizer for him. David starts out with 33% confidence in each of the three options being the one that his friend would like most. He then tries to remember what he knows about his friend's tastes, but he can't really come up with anything that supports one of the options as being a better choice than the others. He concludes his reasoning by being

33% confident that each of the options would be most to his friend's taste, and decides what to order by using a random number generator on his phone. In this example, David's credences remain the same the whole time, but only when they become terminal attitudes, i.e., when David has satisfied himself that he has tried to figure out to the best of his knowledge what his friend would like, David is ready to act on these credences. Hence, it is not due a change in his credences, but due to a change in their status from transitional to terminal, that Action comes to apply to them. David wouldn't have decided by generating a random number at the beginning of his reasoning what to order, because at that time, it was still an open possibility for him that he might come up with a reason to favor one of the appetizers over the others. We can easily come up with lots of cases that work similarly, which is further evidence for the claim that, rather than being essential features of belief, Action, Affirmation and Premising are key features of terminal attitudes.

The second argument for the claim that Affirmation, Premising and Action are best thought of as applying to terminal attitudes, rather than beliefs, relies on contrasting belief with nearby categorical attitudes such as acceptance, hypothesis, and supposition. If the attitudes our agents have in the relevant examples are categorical attitudes, but not transitional beliefs, as option 2 claims above, then we should see this in the features possessed by these attitudes. Of course, those attitudes could be categorical attitudes that are of some *sui generis* kind, rather than a species of a known attitudes, like acceptance, hypothesis, or supposition. Yet, it seems methodologically preferable to first consider attitudes that are already part of our usual taxonomy.

Interestingly, the features that we found to be continuous between the transitional attitudes in our examples and terminal beliefs don't apply to acceptance, hypothesis, and supposition. First, take Involuntariness: It's widely accepted that these attitudes are under our voluntary control. We are free to accept, hypothesize, and suppose whatever we like, regardless of what our evidence suggests at that point in time. Second, consider Feeling: since we can accept, hypothesize and suppose whatever we like, there is no specific feeling of truth associated with this attitude. Of course, we might sometimes suppose, accept, or hypothesize something that we also feel very confident in. But this confidence is not due to the fact that we have one of those three attitudes towards the claim in question. Rather, this confidence is due to how strongly we believe the content, which we can then *also* suppose, accept, or hypothesize. Third, consider Surprise: There's nothing particularly surprising about finding out that one of our hypotheses, acceptances or suppositions is false. If we were to feel surprised, it would be in virtue of how confident we felt about the claim in question, not in virtue of the fact that we accepted, supposed, or hypothesized it. Hence, with regard to these three features, the attitudes under consideration seem more like beliefs than like acceptances, hypotheses or suppositions. This speaks in favor of accepting option 1, which categorizes them as transitional beliefs, instead of option 2.

The remaining features from Alston's list (Inference, Action, Premising, Affirmation) are a better fit for hypothesis, acceptance and supposition, with the caveat that they tend to only apply within a particular context of reasoning. An agent who adopts such an attitude tends to only do so within a circumscribed territory, but as long as they are operating within this territory, they will use the relevant claim as a premise in reasoning, as a basis for action, and so on. But the fact that these features apply to hypothesis, acceptance and supposition is evidence against both the claim

that they are key features of belief, and evidence against the view that the attitudes in our examples are hypotheses, acceptances or suppositions. It is evidence against the former claim because, if these features were distinctive characteristics of belief, we shouldn't expect them to also apply to a variety of other attitudes. Rather, it is evidence for the claim that they are features of terminal attitudes, since, based on traditional categorization schemes, hypothesis, acceptance and supposition would also have been thought of as varieties of terminal attitudes (insofar as the distinction between terminal and transitional attitudes can be meaningfully applied to them).¹³

Further, the fact that Inference, Action, Premising, Affirmation apply to hypothesis, acceptance and supposition is evidence against the claim that the attitudes in our examples belong to one of those kinds. This is because the attitudes in our examples don't exhibit Action, Premising, and Affirmation, so it's unfitting to categorize them as attitudes that have these features.

In sum, the feature approach to characterizing beliefs does not tell against the view that agents can have transitional beliefs, once we distinguish which features are really features of belief, and which are features of terminal attitudes more generally. Moreover, it doesn't favor categorizing the attitudes in question as hypotheses, suppositions or acceptances instead. While some philosophers favor this feature-bundle approach to characterizing belief, others try to instead isolate the distinctive features of belief by identifying what the aim of belief is. I will turn to this approach, and what it entails for our question about the existence of transitional beliefs, in the next section.

3.2 Belief as Having a Characteristic Aim

Some philosophers take as their starting point for theorizing about belief the metaphorical claim that belief has a characteristic aim, which is usually taken to be truth. While the metaphor is explicated in different ways by different philosophers, a commonality between the views is that belief is taken to be subject to a particular correctness standard or truth norm, hence, beliefs that aren't true are normatively defective. Different explanations are given for why belief is subject to this characteristic norm: some argue for a constitutivist approach, according to which it is part of the nature of belief to be governed by a norm according to which beliefs are correct if and only if they are true (e.g., Wedgwood 2002). Others have argued that the correctness standard for belief is generated by a particular normative power of the believer, who subjects a claim to this standard by believing it (e.g., Singh 2023).

Once this correctness standard has been established, it is used to explain a variety of further features of belief. For example, it can be used to distinguish beliefs from other attitudes that involve treating a claim as true, such as hypothesis, acceptance, or supposition. While there are norms governing when those latter attitudes can be rationally adopted, they are not governed by some sort of essential truth norm, since there is nothing inherently wrong with accepting, hypothesizing or supposing something that is false (see, e.g., Shah & Velleman 2005). Further, the truth norm is used to explain why beliefs should be responsive to evidence. Since human thinkers don't have

¹³ It's not obvious to me that it makes sense to speak of transitional versions of suppositions, acceptances or hypotheses, but I will leave a proper discussion of this for another occasion.

access to the truth directly, they can't get their beliefs to somehow directly match the truth. Rather, the best means available to them is to match their beliefs to their evidence, since their evidence is an accessible guide to the truth for them.

Like with other theories of belief, aim-theories of belief were conceived of with terminal beliefs in mind. Yet, the ways in which they conceive of belief naturally extend to transitional beliefs. While transitional beliefs are not an agent's considered opinion about the matter they are deliberating about, they are still regulated by a truth norm. The reason why the agent engages in deliberation is because they want to find out the truth about the question under consideration, or at least the next best thing: what their information tells them about this question. If an agent forms a false transitional belief during this process, there is a clear sense in which this belief is incorrect. For example, Tilda's transitional belief that the formula is not a tautology is not true, and it is defective in virtue of that.

Further, the correctness standard is used to explain why beliefs should be responsive to the agent's evidence. We see this norm of evidence-responsiveness in transitional beliefs just like in terminal beliefs, the only difference being that terminal beliefs are subject to a total evidence standard, whereas transitional beliefs are supposed to reflect the agent's insight into their evidence that they have reached at a given point in their deliberation. Yet, both transitional beliefs and terminal beliefs are subject to the same rational pressure to be adjusted when the agent acquires further information or insights that bear on their truth. Hence, views of belief according to which the truth aim (however spelled out) is an essential feature of belief present no obstacle to positing the existence of transitional beliefs, and to attributing them to the agents in our examples.

3.3 Belief as a Settling Attitude

The last account of belief I will consider characterizes belief via its role in inquiry. It has recently been argued that belief is special in that it is an inquiry-settling attitude. If this is the case, then it's either irrational to believe that p and to keep inquiring into whether p , or even impossible to do so.¹⁴ Friedman (2019) argues for the normative version of this claim, according to which there is a coherence norm which forbids believing a complete answer to some question while at the same genuinely inquiring into the answer to this question. Lee (2023) argues for a descriptive interpretation, according to which it is impossible to wonder whether p and occurrently believe that p at the same time. According to both Lee and Friedman, this distinguishes beliefs from credences and agnostic attitudes, since it is neither incoherent nor impossible to hold one of those attitudes towards a claim p while continuing to inquire into whether p . Further, this also distinguishes belief from acceptance, hypothesis, and supposition, since those attitudes are compatible with the possibility and rational permissibility of inquiry.

As before, the first important thing to observe is that both Lee and Friedman have terminal beliefs in mind in making their arguments. It makes sense that a terminal belief is incompatible, either in a descriptive sense or in a normative sense, with further inquiry (though see, e.g., Falbo

¹⁴ Other proponents of this type of view include Hieronymi (2005, 2008), Fantl & McGrath (2009) and Wedgwood (2012).

2023 for a different perspective). This is because a belief represents one of the possible answers to the question under consideration as true, and if the belief is terminal, then, by the agent's own lights, it is supported by sufficiently good reasoning that no further deliberation is needed. Hence, there is no further uncertainty to resolve, and no potential deficiency in one's reasoning to improve upon. This means that the agent has no good reason to keep inquiring into p , or so the line of reasoning goes on the view that belief is a settling attitude.

Consequently, an attitude can fail to be an inquiry- settling attitude either in virtue of being a terminal attitude that leaves open some uncertainty, like terminal credences or terminal agnostic attitudes, or in virtue of being not terminal. While a transitional belief represents one answer to the question under consideration as the correct one, the agent has not arrived at this attitude via a reasoning process that they deem sufficiently good, either because they think they have more evidence to consider, or because they want to deliberate more deeply or carefully about the matter. Hence, an agent who has this kind of attitude is not in a position to consider the question at hand to be settled; rather, they have good reason to keep inquiring. This means that the view that beliefs are settling attitudes is not plausible if we interpret it as applying to both transitional and terminal beliefs. Beliefs are only suitable for settling an inquiry when they have the status of being terminal, because that endows them with the needed features for ending inquiry. Transitional beliefs are not settling attitudes. While being a belief of some kind might be a necessary condition for an attitude to be a settling attitude, it's not a sufficient condition.

A further consequence of this view is that an agent doesn't need to stop believing something if they want to reopen inquiry. On Friedman's and Lee's views, a rational agent who occurrently (and terminally) believes that p needs to stop believing it and switch to an inquiring attitude towards p if they want to reopen their inquiry into the question at hand. On my view, by contrast, this is not required. If an agent terminally believes that p as an answer to Q , but comes to think that further deliberation about how to answer Q is required, they need not abandon their belief in p , rather, they can switch its status from being terminal to transitional.

There is one further loose end here that arises from Friedman's normative principle DBI, which says that it's incoherent to both believe a complete answer to some question and actively inquire into how to answer this question at the same time. An agent who believes that p , yet keeps trying to find an answer to the question to which p is a complete answer, looks like they might be in a state of epistemic akrasia. Epistemic akrasia is a type of level-incoherence, where the agent holds some first-order doxastic attitude while also having a second-order attitude that calls into question the rationality of their first-order attitude. If an agent keeps inquiring despite already believing an answers to the question at hand, this might be because they don't think their first-order attitude is justified. And this kind of level-incoherence has been deemed objectionable by many epistemologists. In the next section, I will discuss whether these considerations can be developed into a normative challenge for the rational permissibility of holding transitional beliefs.

4. The Challenge from Epistemic Akrasia

In this section, I will address the question of whether there is a normative obstacle to positing transitional beliefs: It seems that those beliefs would inevitably be akratic and thus rationally

objectionable. Epistemic akrasia is standardly characterized as involving an agent having a first-order doxastic attitude towards p , while also thinking that this attitude towards p is rationally defective or not reasonable for them to have. For example, I might believe that my new freezer will be delivered next week, while I also think that I have no good reason to think so, and that my belief is merely based on wishful thinking and unhinged optimism. My attitudes in this case display epistemic akrasia, since my first-order reasoning leads me to have an attitude towards p (the claim that my freezer will be delivered next week), but my higher-order evidence suggests that there is a problem with my first order reasoning and/or attitude.

Many epistemologists think that it is epistemically irrational to be akratic in this way, for several reasons (see, e.g., Horowitz 2014, Titelbaum 2015): first, akratic attitudes exhibit a problematic kind of inter-level tension. Second, this makes akratic attitudes a bad guide for action, since the agent's first- and higher-order attitudes give conflicting guidance (Order a bunch of frozen food for next week! No, don't order frozen food for next week!). Third, it seems like the agent could use their first-order attitude to deduce that their higher-order attitude is a misleading defeater, but this seems like illegitimate bootstrapping. These are not the only problems that have been pointed out with having akratic attitudes, but this will suffice to get a general idea of the problems. Some epistemologists have argued that there are exceptions to the rational impermissibility of having akratic attitudes, but those cases usually involve unusual setups, such as uncertainty about what one's evidence is, cases in which one's evidence is falsity- rather than truth-guiding, and cases in which rationality and accuracy come apart (Williamson 2011, Horowitz 2014, Christensen 2024, though see Lasonen-Aarnio 2014 and Hawthorne et al. 2021 for a more sympathetic take on the rational permissibility of epistemic akrasia).

When agents find themselves in a situation in which their higher-order and their first-order attitudes conflict (and they are not in one of these exceptional cases), a common recommendation is that they need to resolve the tension between their attitudes by adjusting their first-order attitudes. For example, if I have good reason to think that my belief that the new freezer will be delivered next week is not rational, I should reduce my confidence in this claim and/or adopt an agnostic attitude about it (see, e.g., Christensen 2007).

This pretty common view about epistemic akrasia might be thought to present a normative challenge for the view that rational agents can have transitional beliefs: When an agent's first-order reasoning generates a transitional belief, then, by definition, they don't think that they are in a position (yet) to turn the belief into a terminal attitude. This is because they think the belief is not (yet) sufficiently supported by good reasoning. Hence, their attitudes exhibit the characteristic inter-level tension between a first-order attitude that has been generated by first-order reasoning, and a higher-order attitude that deems the first-order attitude and/or reasoning somehow rationally deficient.

If we then apply the usual level-merging story, we get the result that in order to be rational, agents should reduce their confidence in the relevant first-order claims, which means that they need to adopt transitional credences or transitional agnostic attitudes in order to alleviate the inter-level tension. If this is the right response, then it favors adopting option 3, from above, rather than option 1:

Option 1: It is possible for agents to have rational transitional beliefs, and those are the attitudes we should ascribe to David and Tilda.

Option 3: It is not possible for agents to have rational transitional beliefs, and the attitudes agents can rationally have in cases like David's and Tilda's are credences and suspensions.

Hence, according to this line of reasoning, David and Tilda can't rationally have transitional beliefs that the dinner bill is \$36.50 per person, and that the formula is not a tautology, respectively. Rather, at the relevant intermediate stage of their reasoning, they must factor their higher-order doubt about the correctness of their first-order attitudes into those, leaving them with non-extreme credences and/or agnostic attitudes towards these claims.

I will now argue that this line of reasoning is mistaken, because the arguments for the irrationality of akrasia within an agent's terminal attitudes don't carry over to akrasia within their transitional attitudes. The first problem lies with the claim that agents need to level-merge because their higher-order attitudes indicate a problem with their first-order transitional attitudes. Level-merging would be a good idea if it somehow produced attitudes that are superior to the first-order transitional attitudes that the agent had to begin with. But it doesn't. Here's why: The reason why the agent's transitional attitudes are deemed deficient is that they are not the result of a sufficiently good reasoning process about the agent's total truth-relevant evidence. But neither are the level-merged attitudes. They are no better than the agent's unmerged first-order attitudes at meeting the standard of having been produced by a sufficiently good reasoning process about the agent's total truth-relevant evidence. In a sense, they are even worse, because they don't even clearly indicate what the results of the agent's truth-relevant reasoning are up to that point, since they are a mix of first-order reasoning and higher-order uncertainty.

We can see the unhelpfulness of level-merging more clearly if we distinguish two standards by which we can evaluate transitional attitudes. One of these is the standard we usually hold terminal attitudes to, and transitional attitudes that reflect a partially completed deliberation fail this standard by design, regardless of whether the agent level-splits or level-merges (which is why the agent has not promoted them to being terminal!). The other standard is one that evaluates whether a transitional attitude is a good reflection of the insights produced by the agent's truth-relevant reasoning up to that point.¹⁵ This is the standard that properly applies to transitional attitudes, given their function in reasoning (Staffel 2021, 2023). The agents' unmerged transitional attitudes in our examples meet this standard, but the level-merged ones don't. Since level-merging does not constitute an improvement in the sense of getting the agent closer to their goal of figuring out what their truth-relevant evidence supports with regard to the question at hand, there's no reason to think level-merging is beneficial in this case.

¹⁵ To be precise, when a transitional attitude is about to be promoted to being a terminal attitude, it should no longer fail the standard of rationality that applies to terminal attitudes. But this edge case is not relevant to my main line of argument here.

In fact, once we think more closely about the mechanics of complex deliberation, it becomes apparent that level-*splitting* is actually beneficial at the transitional stages of reasoning. Recall that the agent's goal in deliberating is to figure out how the information they possess bears on answering the question under consideration. The best method for achieving this goal is to carefully reason about their truth-relevant evidence until they have reached a conclusion via a sufficiently good deliberation process. While the process is underway, the agent's metacognitive monitoring processes will indicate whether the process has reached an appropriate stage to be terminated. If everything goes well, there will be harmony between the agent's first- and higher-order attitudes at the concluding stage, since the agent's higher-order attitudes will approve of the agent's first-order conclusion.

This process would not be improved upon if the agent constantly factored their higher-order doubts about their first-order transitional attitudes into those attitudes at all the preliminary stages of their deliberation. Constantly changing one's first-order attitudes in order to incorporate higher-order information about the quality of one's reasoning makes reasoning harder to keep track of and is not conducive to the goal of determining how one's truth-relevant evidence bears on answering the question at hand. The cognitive load of doing these updates would be much higher than just reasoning about the truth-relevant evidence, and the agent would ultimately have to disentangle their first- and higher-order evidence anyways, since they are only interested in how their *truth-relevant* evidence bears on the question under consideration. Hence, level-merging does not just fail to improve the rationality of the agent's transitional attitudes, it actually makes achieving the goal of their deliberation more difficult. Hence, a form of *transitional akrasia*, in which the agent forms transitional attitudes based on their first-order reasoning, and keeps these separate from the higher-order monitoring attitudes that evaluate them, is beneficial given the mechanics and aims of complex deliberation.

Lastly, consider the standard reasons for which akrasia in one's terminal attitudes is considered problematic, such as the conflicting action guidance issued by akratic attitudes: since transitional attitudes typically don't guide action, precisely because they are not the agent's considered opinion on the matter, we don't have to worry about them issuing problematic recommendations for how to act. Another worry is that akratic attitudes allow a problematic form of bootstrapping, in which the agent gets to dismiss their higher-order doubts based on their first-order attitudes. This is not a problem for cases of *transitional akrasia*. Since transitional attitudes are not licensed for unrestricted use as premises in reasoning, they are not available as premises in inferences that would lead the agent to dismiss their higher-order attitudes that are the precise grounds on which the first-order attitude is deemed transitional.

To sum up: Arguments against the rational permissibility of epistemic akrasia seemed initially to also tell against the view that agents can have rational transitional beliefs. It seemed like arguments in favor of level-merging might rationally require reasoners to only adopt transitional non-extreme credences and agnostic attitudes, which reflect the reasoners' higher-order doubts about the correctness of the first-order attitudes. Yet, those arguments turned out to be unsuccessful upon closer inspection. I argued that a reasoner's deliberative aims are in fact interfered with rather than promoted if the agent tries to avoid *transitional akrasia* by level-merging their first- and higher-

order attitudes while deliberating. Moreover, standard worries about akrasia in one's terminal attitudes don't carry over to *transitional akrasia*, because transitional attitudes are not typically used unrestrictedly as bases for action, or as premises in further reasoning. Hence, while it initially seemed that challenge from akrasia might push us towards accepting option 3 instead of option 1 from our menu of available views, further inquiry revealed that this is not so. Notice that my arguments in this section don't establish that level-merging in cases like David's and Tilda's is always rationally forbidden. Rather, they establish that level-splitting is both rationally permissible and advantageous for facilitating efficient deliberation. Hence, they show that it is rationally permissible to have transitional beliefs, which is what I was hoping to demonstrate.

Conclusion

The question I set out to answer in this paper was: Is it ever rationally permissible to *transitionally* believe something before one has concluded one's deliberation, i.e., before one has settled on an answer to Q? I argued that the answer to this question is positive: In cases like Dinner Bill and Logic Problem, it is rational for the agents to adopt transitional beliefs while they are still in the process of deliberation. I considered two challenges to adopting this view: The challenge from the nature of belief, and the challenge from epistemic akrasia. Regarding the former, I argued that common views of the nature of belief are really accounts of terminal beliefs. Once we disentangle which of the features posited by these views apply to beliefs specifically, and which ones apply to terminal attitudes, we can see that these views are compatible with positing transitional beliefs. Further, the discussion of this challenge also brought out that the attitudes in questions are better characterized as beliefs, rather than alternative categorical attitudes such as acceptances, hypotheses or suppositions.

Regarding the challenge from epistemic akrasia, I showed that common arguments against the possibility of rational akrasia in one's terminal attitudes and in favor of level-merging don't carry over to akratic transitional attitudes. Rather, level-splitting seems rationally beneficial while an agent is in the process of deliberating. As a result, there is no normative pressure to abandon transitional beliefs in favor of transitional credences or transitional agnostic attitudes.

If I am right about all of this, then we must revise our existing theories of (rational) belief in a number of ways. First, we need to update our descriptive theories of belief: the popular view according to which beliefs can be characterized by bundles of dispositions typically associated with them must be revised, so as to disentangle which of the listed dispositions are really features of belief, and which ones are features of terminal attitudes. Secondly, theories that see beliefs as settling attitudes require revision as well. These views are at best correct about terminal beliefs, but not about beliefs more generally. Moreover, counter to what these theories claim, it is not necessary for agents to abandon a belief and adopt a non-extreme credence or agnostic attitude when they reopen their inquiry into a question. Rather, it can be rational to keep the belief, merely changing its status from terminal to transitional. Thirdly, my arguments have significant implications for the debate about the rational permissibility of epistemic akrasia. It is typically assumed that having akratic doxastic attitudes is epistemically bad, and that cases that are exceptions to this are fairly weird. On my view, having epistemically akratic attitudes during ongoing deliberations is

epistemically rational, practically efficient, and commonplace. Hence, we have to revise our views of epistemic akrasia, either by recognizing its benefits for complex deliberation, or by narrowing the scope of the concept so it doesn't apply to transitional attitudes.

Acknowledgements

I would like to thank Bret Donnelly, Tyler Porter, Yurong Zhu, Amir Ajalloeian, and Jonathyn Zapf for helpful comments and discussion.

References

- Ackerman, Rakefet. 2014. The Diminishing Criterion Model for Metacognitive Regulation of Time Investment. *Journal of Experimental Psychology: General* 143 (3), 1349-1368.
- Ackerman, Rakefet & Thompson, Valerie A. 2017. Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences* 21 (8), 607-617.
- Alston, William. 1996. Belief, Acceptance and Religious Faith, in Jordan and Howard-Synder (eds.), *Faith, Freedom and Rationality*, Rowman and Littlefield Publishers, 3-27.
- Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Christensen, David. 2007. Does Murphy's law apply in epistemology? *Oxford Studies in Epistemology* 2, 3-31.
- Christensen, David. 2024. Epistemic Akrasia: No Apology Required. *Noûs* 58 (1), 54-76.
- Cohen, Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.
- Falbo, Arianna. 2023. Inquiring Minds Want to Improve. *Australasian Journal of Philosophy* 101 (2), 298-312.
- Fantl, Jeremy & McGrath, Matthew. 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Friedman, Jane. 2013. Rational Agnosticism and Degrees of Belief. *Oxford Studies in Epistemology* 4 (57), 57-81.
- Friedman, Jane. 2019. Inquiry and Belief. *Noûs* 53 (2), 296-315.
- Good, I.J. 1967. On the Principal of Total Evidence. *The British Journal for the Philosophy of Science* 17 (4), 319-321.
- Hawthorne, John, Isaacs, Yoaav & Lasonen-Aarnio, Maria. 2021. The Rationality of Epistemic Akrasia. *Philosophical Perspectives* 35 (1), 206-228.
- Hieronymi, Pamela. 2005. The Wrong Kind of Reason. *Journal of Philosophy* 102 (9), 437-457.
- Hieronymi, Pamela. 2008. Responsibility for Believing. *Synthese* 161 (3), 357-373.
- Horowitz, Sophie. 2014. Epistemic Akrasia. *Noûs* 48 (4), 718-744.
- Lasonen-Aarnio, Maria. 2014. Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88 (2), 314-345.
- Lee, Wooram. 2023. Belief and Settledness. *Inquiry* 66 (6), 1094-1119.
- Schwitzgebel, Eric. 2002. A phenomenal, dispositional account of belief. *Noûs* 36, 249-275.
- Shah, Nishi & Velleman, David. 2005. Doxastic Deliberation. *The Philosophical Review* 114 (4), 497-534.
- Singh, Keshav. 2023. Belief as Commitment to the Truth. forthcoming in: Eric Schwitzgebel & Jonathan Jong, *What is Belief?*, Oxford: Oxford University Press.

- Smithies, Declan. 2023. Belief as a Feeling of Conviction. forthcoming in: Eric Schwitzgebel & Jonathan Jong, *What is Belief?*, Oxford: Oxford University Press.
- Staffel, Julia. 2019. Credences and Suspended Judgments as Transitional Attitudes. *Philosophical Issues* 29 (1), 281-294.
- Staffel, Julia. 2021. Pro Tem Rationality. *Philosophical Perspectives* 35 (1), 383-403.
- Staffel, Julia. 2023. Transitional Attitudes and the Unmooring View of Higher-Order Evidence. *Noûs* 57 (1), 238-260.
- Staffel, Julia. 2024a. *Unfinished Business: Rational Attitudes in Reasoning*. Manuscript.
- Staffel, Julia. 2024b. Suspension in Inquiry. Manuscript.
- Steup, Matthias. 2017. Believing Intentionally. *Synthese* 194 (8), 2673-2694.
- Thompson, Valerie A. & Johnson, Stephen C. 2014. Conflict, Metacognition, and Analytic Thinking. *Thinking and Reasoning* 20 (2), 215-244.
- Titelbaum, Michael. 2015. Rationality's Fixed Point. *Oxford Studies in Epistemology* 5, 253-294.
- Wedgwood, Ralph. 2002. The Aim of Belief. *Philosophical Perspectives* 16, 267-297.
- Wedgwood, Ralph. 2012. Outright Belief. *Dialectica* 66 (3), 309-329.
- Weisberg, Jonathan. 2020. Belief in Psyontology. *Philosophers' Imprint* 20 (11), 1-27.
- Willard-Kyle, Christopher. 2022. Valuable Ignorance: Delayed Epistemic Gratification. *Philosophical Studies*, online first.
- Williamson, Timothy. 2011. Improbable Knowing. In: Trent Dougherty (ed.), *Evidentialism and its Discontents*, Oxford University Press, 147-164.