Canadian Journal of Philosophy (2026), 56: e40, 1–17
doi:10.1017/can.2026.10057

SYMPOSIUM

Humble Reasoning: When can I Stop?

Julia Staffel

CU Boulder: University of Colorado Boulder, USA
Email: staffel.julia@gmail.com

Abstract
One way of characterizing what makes someone a good reasoner is to appeal to intellectual virtues, such as curiosity, fair-mindedness, or epistemic humility. My aim in this paper is to show that explaining how the virtue of humility should manifest itself in complex reasoning is more difficult than one might think. A very natural view of what intellectually humble deliberation looks like is problematic, because it leads to an infinite regress. I will explore whether and in which way this regress is vicious, and how our answers to these questions can lead us to a better account of how humility can inform stopping rules for good reasoning.

Keywords: reasoning; humility; virtue; regress; justification; rationality; inference

1. Introduction

One way of characterizing what makes someone a good reasoner is to appeal to intellectual virtues, such as curiosity, fair-mindedness, or epistemic humility. Being epistemically (or intellectually) humble is thought to promote open-mindedness and counteract dogmatism in our intellectual endeavors, thereby promoting the acquisition of knowledge and other epistemic goods. But what exactly is involved in manifesting this virtue that lets one reap those benefits? The question I will explore in this essay concerns a specific application of epistemic humility: How should a humble reasoner deliberate? I will argue that a natural view of what intellectually humble deliberation looks like cannot be correct, because it leads to a problematic regress. To avoid it, we must ignore some of our imperfections. Humble reasoning thus requires a certain amount of immodesty, or, in more positive terms, self-trust.

In what follows, I will first survey some recent accounts of intellectual humility. I will then sketch an empirically informed account of the way human reasoners engage in complex deliberation, and explain how accounts of humility can help us answer the question of how a humble reasoner should deliberate. The resulting view is one on which reasoners should be sensitive to the quality of their own reasoning and try to ensure that it is sufficiently good, yet they should avoid striving for absolute certainty that they have reasoned correctly. This picture of humble deliberation gives rise to a more specific question about residual error possibilities: once a humble reasoner has finished deliberating, what should they do about their residual uncertainty about whether they have reasoned well? I argue that a natural answer to this question is given by the *adjustment view*, which is a natural consequence of any account that always requires *level-merging*, that is, that reasoners should adjust their conclusions in light of evidence of their own fallibility. However, the adjustment view cannot be correct, because it leads to a vicious regress. The regress arises because any time an agent engages in reasoning, which can include adjusting the conclusions of their previous reasoning, they are left with a residual uncertainty that they might have made an error, requiring further adjustments of their attitudes.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

I spend the remainder of the paper explaining why this regress is problematic and exploring what kind of solution we should seek for it. Our best way out, I conclude, is to embrace a *benchmark view* on which reasoners may simply disregard residual error possibilities, as long as they are insignificant enough. On this view, humility manifests itself in setting a good benchmark, which strikes the right balance between self-trust and self-scrutiny. The resulting view fits well with the most plausible theories of intellectual humility, and it nicely parallels popular fallibilist views of belief and knowledge.

2. Humility in Reasoning

Intellectual humility is supposed to be an epistemic virtue that leads to greater knowledge and understanding, invites learning and correction of mistakes, and avoids dogmatism. In this section, I will survey various accounts of intellectual humility in preparation for answering the question of how it can help us reason well.

Intellectual humility is a multifaceted concept, with different accounts highlighting different interesting aspects of it. Some accounts emphasize that humility requires setting aside one's ego when one is pursuing the truth. Callahan offers an account of intellectual humility that prioritizes the humble thinker's ability to pursue the truth free from ego-related distractions. She argues that "vicious intellectual pride always disposes one to distraction; the viciously intellectually proud agent is prone to distraction by their believing self or ego in ways that occlude or interfere with effective intellectual activities" (Callahan, 2024, p. 321). In his survey article on recent work on intellectual humility, Ballantyne likewise emphasizes the connection between humility and an undistorted pursuit of truth:

> "IH is a way for us to manage information that's relevant to our pursuit of truth and avoidance of error. The essence of IH, then, is not a matter of how we react to other people, but how we react to information relevant to our inquiry. It is about how we respond to evidence concerning reality. IH involves private mental states and processes, not publicly observable states or interpersonal processes. […] IH 'down-regulates' egoic and egoistic motives in favor of *reality-orientedness*." (Ballantyne, 2023, p. 5)

This means that a failure of humility in deliberation has two main effects. First, an agent who is distracted from the pursuit of truth by their own ego might have an inflated sense of their own skill and level of performance. They think overly highly of their own capacities as a reasoner, which leads them to be overconfident in the quality of their deliberations. For example, someone might have an overly inflated sense of how good they are at mental math, leading them to scoff at the idea of double-checking their calculations. As a result, they often fail to spot their own errors and accept incorrect conclusions. The second way in which someone's ego might interfere with their deliberation is by making them have too little confidence in their own performance. A thinker who is underconfident in their skills as a reasoner might waste time on excessive double-checking, avoid making up their mind, or refrain from acting on their conclusions (see also Roberts & Wood, 2003, 2007, ch. 9).

This suggests that a key component of humility in reasoning is the capacity to accurately assess one's own skills and the quality of one's performance. This aspect of intellectual humility is also emphasized in the literature on the topic. Tanesini (2018) argues that a necessary condition for intellectual humility is to have at least a relatively accurate view of one's cognitive abilities and limitations (see also Zagzebski, 1996, p. 220, Church, 2016, Whitcomb et al., 2017). Hazlett (2012) suggests that an intellectually humble person must accurately assess the epistemic quality of their first-order attitudes. Sosa (2011) concurs—he argues that acquiring the highest level of knowledge, *knowing full well*, requires that the agent uses meta-knowledge of the quality of their first-order belief-forming methods to guide their reasoning. He says, "A performance is *fully* apt only if its first-order aptness

derives sufficiently from the agent's assessment, albeit implicit, of his chances of success (and, correlatively, of the risk of failure)" (Sosa, 2011, p. 11).

While this gives us a general sense of how humility contributes to good reasoning, it is difficult to say anything more specific unless we take a more detailed look at how humans reason, and which aspects of their reasoning performance might be altered or improved by the presence or absence of humility. Hence, in the next section, I will sketch an empirically informed view of human deliberation, which will help us do so.

3. Metacognition and Complex Deliberation

When we ask whether a person is intellectually virtuous (or vicious), we ask about a particular aspect of their cognitive profile, namely, whether they possess certain kinds of excellences and associated skills and capacities to greater or lesser extents.[1] In the epistemic realm, we are thus asking about aspects of people's pursuit of truth that they have at least some degree of control over, and that they could execute more or less excellently. A central locus in which intellectual virtues manifest themselves is inquiry, which includes activities such as deliberation and gathering new information.[2]

Let us first get clear on what "deliberation" means in this context. Deliberation is different from inquiry or learning new information. Deliberation or reasoning (I will use these terms interchangeably) involves drawing out the consequences of information one already possesses. It is thus a narrower notion than learning or inquiry, which additionally includes acquiring new empirical information, formulating new research questions, and related activities. The question of how a humble thinker should *inquire* thus includes, as a sub-question, how humility should manifest itself in *deliberation*, but is not exhausted by it.

In characterizing deliberation or reasoning, we can further distinguish between simple and complex reasoning. Many inferences we draw are very simple and are conducted automatically, such as basic modus ponens inferences, cases of "and" elimination, and so forth. There's broad agreement in the philosophical literature that we have basic justification for drawing these types of inferences (although it is controversial why this is so, and how to delineate the class of inferences for which we have this basic justification, see Dogramaci, 2012, Schechter, 2013, 2019). For these simple inferences, it does not make much sense to ask how a humble reasoner should draw them, as there aren't really different ways of drawing them.

The question of how to reason humbly is more relevant for complex deliberations in which the thinker employs an extended cognitive process to answer a question or solve a problem. In these cases, the agent typically has some amount of choice and control over how to deliberate; hence, the question of how to reason humbly is intelligible. This is not to say, however, that complex reasoning must be entirely conscious and cannot be executed automatically for the question of how to reason humbly to make sense. Our habits and dispositions can also exhibit humility to a greater or lesser degree, and we can influence them by cultivating better (or worse) intellectual practices over time. In short, I will focus on processes of complex deliberation, in which the reasoner figures out what to think about a question of interest in light of the information they already possess.

To better understand how human reasoners carry out complex deliberations, we can turn to the empirical literature on metacognition. One strand of the psychology literature on problem-solving targets the question of how reasoners carry out complex reasoning strategies. This research is not aimed at describing how we solve particular types of problems, but at a more general account of the cognitive mechanisms involved. Generally speaking, there is agreement that complex problem-

[1]For a discussion of the relationship between intellectual virtues and skills, see Zagzebski (1996), p. 115).

[2]Forming beliefs or credences, by contrast, is largely involuntary, and thus not a central manifestation of one's intellectual virtue or vice.

solving employs both first-order and higher-order processes. The first-order process is the strategy that the agent employs to solve a given problem, while higher-order processes are deployed simultaneously to monitor and control the execution of the first-order process. The reasoner uses these higher-order metacognitive processes to determine whether the problem-solving strategy is being carried out correctly or needs to be adjusted, and whether it has delivered an acceptable conclusion (see, e.g., Ackerman & Thompson, 2017; Thompson & Johnson, 2014).

Refining this general picture, Ackerman (2014) proposes a more specific model of how agents decide when to terminate a deliberation. According to the *Diminishing Criterion Model*, reasoners set a target level of confidence that they have reasoned correctly, which determines an appropriate stopping point for their deliberation. However, this level of confidence is not necessarily static and can change over time. If an agent realizes after starting their reasoning process that the initially set confidence target cannot be reached, they will either give up or they will lower their target level of confidence and settle on an answer that passes this adjusted threshold. Whether the agent gives up or settles for a lower quality answer depends on various factors, such as the exact nature of the problem to be solved, how much time they have, and how important it is to them to come up with an answer.[3]

To suit our purposes, we can formulate a simple model of complex deliberation based on these empirical insights. I will focus on theoretical reasoning rather than practical reasoning, although it would be interesting to apply the subsequent discussion to the justification of intentions as opposed to beliefs. The simple model assumes that the goal of deliberation is to answer some question Q. Since the agent does not typically have direct access to the answer to Q, they must use the best method available to them for answering Q, which is to determine how their truth-relevant evidence bears on answering Q.[4] The agent implements this by executing various first-order reasoning strategies that allow them to evaluate the evidence's significance for Q, where these processes are monitored and controlled by metacognitive processes. Once the agent's metacognitive monitoring indicates that the reasoning has been completed correctly, the agent terminates the process and adopts the resulting conclusion as the answer to Q. This answer can be a belief, but it could also be an attitude of uncertainty or suspension, depending on the agent's evidence. In an instance of good reasoning, this process terminates in a conclusion that the agent is justified in holding based on evidence.

To illustrate this, suppose someone is trying to solve a difficult or counterintuitive math problem, such as the bat and ball problem, which is one of the questions on the *Cognitive Reflection Test* (Frederick, 2005): A bat and a ball cost $1.10 together, and the bat costs $1 more than the ball. How much does each cost? An intuitive answer that jumps out at most people is that the ball costs $0.10. But of course, that means the bat would have to cost $1.10, which cannot be right, since then they would cost $1.20 together. An agent might at first be drawn to this answer, but at the same time have low confidence in the quality of their initial reasoning, perhaps because they know this is a trick question or because they realize the answer violates the initial stipulation about the combined price. This inter-level tension drives the agent to continue their reasoning. Once they see that the intuitive answer of $0.10 for the price of the ball is too high, adjusting the number down a little and checking the sum makes it easy to figure out that the ball must cost $0.05 and the bat $1.05. Having checked the math, the agent is now well positioned to conclude that their metacognitive attitudes are judged to be supported by good enough reasoning.[5]

[3] Ackerman develops her model in contrast to the *Discrepancy Reduction Model*, which assumes that the agent sets a specific target level of confidence in the quality of their reasoning at the beginning of a deliberation, which remains unchanged. Ackerman argues convincingly that the Discrepancy Reduction Model makes implausible predictions because it assumes a static confidence target. For detailed discussion, see Ackerman (2014), Ackerman et al. (2023), Ackerman and Levontin (2023), Law et al. (2022), and Lee et al. (2023).

[4] By *truth-relevant evidence* I mean any evidence the agent possesses that bears on the correct answer to Q, excluding any kind of evidence that bears on the quality of agent's reasoning about Q without directly speaking to the truth of Q.

[5] Of course, if someone trusts their intuitive feeling of rightness and endorses the $0.10 answer, they will accept this answer as the conclusion, as predicted by the model. However, their conclusion would not be justified by their evidence in this case.

Given this model of complex deliberation, the question of how being humble might contribute to good reasoning mainly affects two of its elements. The first element is the choice of strategy for evaluating one's evidence to determine its significance for answering *Q*. Which strategies are best to choose is difficult to answer at a general level, since different kinds of reasoning problems require very different kinds of approaches. In what follows, I will presuppose that the agent has selected a suitable strategy to solve a given problem.

The second way in which humility can affect the quality of one's reasoning is via the metacognitive evaluation and control of the first-order processes. The virtuous agent can use these metacognitive tools to gain accurate assessments of the quality of their first-order reasoning processes. Then, they can use this information to decide how to proceed in their deliberations. Note that this language of "assessing," "guiding," and "controlling" can wrongly suggest that all of these activities are executed with conscious awareness of the agent. Of course, metacognitive awareness and control can sometimes be under the agent's conscious control, but it is also often executed automatically and/or habitually. As Ballantyne (2023) and Tanesini (2018) emphasize, being intellectually humble may or may not involve conscious attitude or deliberation management. Humble dispositions can also be automatic and do not need to involve explicit beliefs or intentions. This is compatible with virtuous deliberation being a matter of competence and skill, as the possession and development of skills usually involves cultivating and relying on good habits and dispositions (Zagzebski, 1996, p. 116).

Suppose, then, that the humble reasoner is skilled at accurately assessing the quality of their own reasoning using suitable metacognitive mechanisms. How confident do they need to be that their reasoning has been completed correctly before they are willing to terminate it and accept its conclusion? If the level they aim for is too low, they would be willing to rely on first-order attitudes that, by their own lights, are likely epistemically defective and do not achieve the goal of pursuing the truth. Hence, a humble reasoner must aim for a high enough confidence in their reasoning quality to judge the resulting conclusion as a successful instance of pursuing the truth.

Yet, at the same time, since epistemic humility requires being realistic about one's capacities, a humble reasoner cannot aim at certainty that their reasoning has been well executed. Given human limitations, being a perfect reasoner is impossible, and hence, an accurate assessment of one's reasoning should take this into account. A reasoner who strives for certainty that their reasoning is error-free is either deluded about their own abilities or is very ineffective at allocating their efforts in the pursuit of truth. Within these bounds of not aiming too high or too low, there is, of course, room for reasoners to flexibly adjust their target level of confidence in the quality of their own reasoning. We saw that Ackerman's *Diminishing Criterion Model* explicitly makes room for an adjustable threshold, and this seems justified in light of how a virtuous reasoner proceeds: In matters of little importance, a lower confidence threshold is appropriate than in matters of life and death. Furthermore, in matters of great difficulty, we might need to accept a higher degree of uncertainty about the correctness of our reasoning than when we are thinking about a straightforward matter. To sum up: Being intellectually humble promotes reaching a justified conclusion of reasoning, because it helps the reasoner to accurately evaluate their own performance, to avoid jumping to conclusions or relying on low-quality reasoning, to not strive for the impossible, and to flexibly adjust their standards.

4. Residual Error Possibilities and a Regress

I now want to focus on a particular aspect of the picture of the humble reasoner we just developed. We said that the humble reasoner evaluates their performance accurately (enough) and does not strive for the impossible, that is, for being certain that their reasoning is error-free.[6] This means that

[6]Being able to evaluate one's own reasoning accurately is arguably a complex skill as well that can be cultivated. It involves being attentive and sensitive to cues that indicate smooth or problematic reasoning, knowing when intuitive judgments are likely untrustworthy, and incorporating higher-order evidence about the quality of one's performance, among other things.

the humble reasoner sets a target level of confidence in their reasoning quality that does not demand perfection, leaving a small residual probability that the conclusion they have reached might be flawed (by being wrong or by being supported by bad reasoning). What should they do about this residual probability that they might have reasoned incorrectly? What we have said so far does not settle this question, but it is natural to think, given our virtue-centric framework, that the thinker should not just ignore or dismiss it. Tanesini suggests, for example, that

> a concern with one's own limitations may generate either a sense of defeatism and resignation, or a desire to address them with a view to improve. In the self-accepting individual, the concern for limitations motivates actions to address these defects. This person seeks to remedy those shortcomings that can be lessened, circumvent those which she cannot change, or when nothing else can be done take her limitations into account in her behavior. (Tanesini, 2018, p. 405)

A more specific proposal in this vein is offered by Christensen (2007), who suggests that reasoners should adjust their first-order confidence in light of the residual higher-order probability that they might be mistaken. He formulates the following three rational ideals:

1. (Logic) An agent's beliefs must respect logic by satisfying some version of probabilistic coherence.
2. (Evidence) An agent's beliefs (at least about logically contingent matters) must be proportioned to the agent's evidence.
3. (Integration) An agent's object-level beliefs must reflect the agent's meta-level beliefs about the reliability of the cognitive processes underlying her object-level beliefs.

While each of the principles seems individually compelling, Christensen explains that they create a dilemma when taken together, because we cannot satisfy these epistemic ideals simultaneously: Suppose someone solves a difficult math problem correctly, arriving at x as the answer. Their truth-relevant evidence supports x. So by (Evidence) & (Logic), they should believe/be fully confident in x. But they realize that there is some nonzero probability that they might have made a mistake. By (Integration), they should slightly lower their confidence that x is the correct answer. But if they do that, they are now holding an attitude that is not supported by their truth-relevant evidence, and that violates probabilistic coherence.

Christensen concludes that there is a tension here: Different epistemic ideals can pull in different directions, and we have to live with that. We should not just reject one of the ideals; instead, rational agents should give some weight to each ideal. It follows that Integration cannot be ignored, and agents should take their own fallibility into consideration when forming their first-order attitudes.[7]

Unfortunately, the problem we run into here is worse than Christensen realizes. If there was just a tension between different epistemic norms or ideals, we could perhaps weigh them by their importance and reach some sort of stable compromise. This would mean that the agent could not fully satisfy all of the ideals, but they could reach a conclusion that optimally balances how well each of the norms is satisfied. The crucial point Christensen's discussion misses is that implementing the adjustments demanded by Integration requires drawing further inferences. Since the reasoner cannot be certain that those further inferences were executed perfectly, they are faced with a residual chance of error yet again, requiring further adjustments, and so on ad infinitum. A regress looms.[8]

[7]For discussion, see also DiPaolo (2019).

[8]Hume discusses this type of regress in his *Treatise of Human Nature* (Hume 1740/2000, Part IV, Section I), but he mistakenly concludes that it would lead agents to assigning zero probability to every claim. See also Meeker and Poston (2010), and Schwitzgebel's discussion here: https://schwitzsplinters.blogspot.com/2022/03/how-to-defeat-higher-order-regress.html.

Let us take this a bit more slowly to see clearly what happens. First, let us articulate more clearly what the humble reasoner should do about the residual probability that they might have made a mistake. As Christensen himself argues, this probability can be small, but it cannot be zero—even ideal agents cannot rule out that they might have made a mistake. Schechter (2013) concurs: He argues that even for applications of basic inference rules, the risk that we have made some kind of error is never zero. Both Tanesini and Christensen say that agents may not simply ignore evidence of their own fallibility: Christensen endorses the integration principle, while Tanesini says that "the concern for limitations motivates actions to address these defects." The most straightforward view that articulates these ideas is the following:

> *Adjustment view:* The reasoner should deliberate about how to answer the question of interest until they are sufficiently confident that they have found the answer through correct reasoning. Any higher-order uncertainty about the correctness of their reasoning becomes part of the total evidence that needs to be accounted for in reaching a justified conclusion.[9]

To be clear, neither Christensen nor Tanesini explicitly endorses this view, but it is a natural extension of a level-merging view that requires that agents always adjust first-order attitudes in light of higher-order evidence of error. Such a level-merging view is popular, and it is hard to see how it could be compatible with rejecting the adjustment view or something closely related.

The adjustment view generates a regress, assuming that the agent is never entirely certain that they have reasoned correctly.[10] This is because adjusting one's first-order attitude in light of one's higher-order uncertainty requires an inference, which itself generates higher-order uncertainty about its own correctness:

1. Deliberate about how to answer the question of interest until you reach a conclusion that meets the threshold of confidence for having reasoned correctly.
2. Adjust your first-order attitude in light of your higher-order residual uncertainty about having reasoned correctly via an inference that meets the threshold of confidence for having reasoned correctly.
3. Adjust the attitude produced by step 2 in light of your higher-order residual uncertainty about having reasoned correctly via an inference that meets the threshold of confidence for having reasoned correctly.
4. Adjust the attitude produced by step 3 in light of your higher-order residual uncertainty about having reasoned correctly via an inference that meets the threshold of confidence for having reasoned correctly.
5. Rinse and repeat…

This regress presents a different problem than the one Christensen identifies, because the thinker cannot arrive at a stable, justified attitude by deciding on how to balance the demands of conflicting ideals. Rather, a stable conclusion eludes the thinker, because every time they adjust their first-order

[9]Many authors have recently defended the claim that reasoners should pay attention to higher-order evidence that they have reasoned incorrectly. However, their views are typically about cases in which the agent believes, or has significant evidence, that their reasoning was problematic, and they are silent about cases in which the probability that one's reasoning was flawed is small.

[10]One might suggest that the regress could be avoided by somehow calibrating one's conclusion to the reliability of the reasoning process employed. For example, if I reach an answer by a reasoning method that I judge to be 95% reliable, then I simply assign 95% confidence to the answer, and no more adjustments are needed. Unfortunately, there are several reasons why this cannot work. One is that simple calibration leads to base-rate fallacies in many cases (see Isaacs, 2019, for discussion). The other is that my confidence that I have applied a method correctly in this case need not match the general reliability of the method, so the residual error probability is not already included in the reliability judgment.

attitude in light of the small chance that their reasoning might have been incorrect, they must do so via another reasoning episode that presents them with the same task all over again: Since they cannot be completely certain that the adjustment of their first-order attitude was a piece of error-free reasoning, another adjustment in light of this small chance of error becomes necessary.

One might wonder whether this regress can be avoided somehow. As an anonymous reviewer points out, empirically informed accounts of rational belief formation do not tend to postulate a separation between arriving at a credence and then adjusting that credence in light of one's own fallibility. Rather, it is assumed that both tasks are somehow executed before a rational prior is reached that integrates both first- and higher-order evidence.

However, to get the regress off the ground, we do not need to rely on a view of reasoning that artificially separates credence formation and adjustment processes. The main problem is that *any* reasoning process carries with it a small risk that it was not carried out correctly, and this risk will generate a regress regardless of how we characterize deliberation processes, assuming it becomes part of the total evidence that needs to be accounted for. For example, suppose we assume instead that agents never endorse a conclusion until they have fully integrated both their first-order evidence and the higher-order evidence about their own fallibility. Whatever attitude the agent reaches, they cannot be certain that whatever process they have used to generate this attitude has been executed correctly. Hence, a residual error possibility remains that must be accounted for before they can conclude, assuming the adjustment view is correct. But any further reasoning steps will again come with a small risk of error, hence the agent can never reach a conclusion that fully incorporates their first-order evidence and the higher-order evidence that they might have made a mistake. The regress takes a slightly different shape here, but it arises nonetheless.

Another suggestion is that the agent could somehow anticipate in advance the risk of error in their reasoning and build it into their conclusions right away. However, this process cannot be more than approximately adequate, because this would mean that the agent is trying to respond to evidence that they anticipate receiving, not evidence they actually have. The fundamental problem is that possessing the evidence that a particular reasoning process might have been flawed requires the agent to have gone through that reasoning. Hence, the agent can only respond to this *possessed* evidence when the reasoning has been completed, which means that anticipatory compliance does not solve the regress problem. (Consider this analogy: a car maker can anticipate the probability that a specific car will have a manufacturing defect. But in order to accurately judge the probability that *this car* has a manufacturing defect, the car has to be made first. For example, if some machine part broke just when this car was going through a crucial manufacturing step, the probability that the car has a defect would be much higher than anticipated.)

In sum, the fundamental problem is that the adjustment view requires agents to respond to any evidence that their reasoning might be flawed. Possessing the evidence that a particular reasoning process might have been flawed requires the agent to have gone through that reasoning. It cannot be anticipated in advance. Since every reasoning process comes with a residual error probability, and responding to evidence requires reasoning, the adjustment view has reasoners forever chase a justified conclusion that accounts for their total evidence. I will now proceed to argue that this regress is not harmless and discuss different ways of avoiding it.

5. Why this Regress is Vicious

It is sometimes taken for granted that any view that generates an infinite regress is thereby bad or wrong. But not every infinite regress is problematic, and those that are can reveal different kinds of flaws in the underlying view. For a harmless regress, consider the claim that "P" entails "P is true." We can plug "P is true" in for "P," generating "P is true is true," and so on. The fact that this can be repeated infinitely many times is not generally considered a problem. Rather, it's a benign feature of how the truth-predicate is commonly understood (Huemer 2016). To argue that the regress

generated by the adjustment view is worrisome, we must identify some underlying problem that the regress brings out.

There are two common failures that regresses put on display that could be reasonably said to apply to the adjustment view: impossibility and explanatory failure (Wieland 2013, Huemer 2016, Cameron 2022). To see how each of them might be exhibited here, recall what the adjustment view was supposed to explain. We started by focusing on the phenomenon of complex deliberation, which can be captured by an empirical model such as the *Diminishing Criterion Model.* We then tried to formulate a criterion for executing this process well by asking how the virtue of humility would guide us to reason, which led us to the *Adjustment View*. We assumed that (1) it is possible for a reasoner to arrive at a justified conclusion while reasoning in a humble way and (2) that the adjustment view correctly captures what humble reasoning consists of (or at least an important aspect of it).

One way to draw out the problem is to point out that the regress reveals an explanatory failure. It shows that the agent cannot reach a stable conclusion that is justified according to the adjustment view. The adjustment view thus fails to meet the explanatory target of giving an account of what is involved in reaching a justified conclusion in deliberation. Since the desired explanation is not delivered, we could then reject our initial assumption that the adjustment view correctly captures how humility leads to good reasoning.

Another way to state the problem is to point out that it's impossible for human reasoners to complete infinite sequences of reasoning. Hence, while the adjustment view gives, in some sense, an account of what it takes to reach a justified conclusion, the requirements it sets out are unattainable for any human thinker, because infinitely many temporally extended reasoning steps are required. We could thus reject the adjustment view as an explication of how humility guides us to reasoning well, because it makes impossible demands.

What both diagnoses have in common is that they retain assumption (1), according to which we have identified a genuine explanatory target, namely, to characterize how a human reasoner can reach a justified conclusion from complex deliberation. But this can also be questioned: Perhaps we were mistaken in assuming that human reasoners (humble or not) can reach a fully justified conclusion as a result of complex deliberation, and the regress brings out this error. There are two ways of spelling this thought out further: One way is to claim that the adjustment view is really an account of what *ideal* rationality requires. The fact that *human* reasoners cannot comply with the adjustment view's demands is thus not necessarily an objection. The second way is to claim that the regress shows that even for ideal reasoners, reaching a justified conclusion of reasoning is impossible, which shows that we were wrong about the possibility of giving an account of such a state altogether. Drawing this lesson from the regress could then be accompanied by giving an error theory of why we mistakenly thought we had identified a genuine target of explanation.

Regardless of which of these diagnoses we favor, they all agree on one thing: that the regress generated by the adjustment view is problematic rather than benign, and that it points to a genuine problem with one of our starting assumptions, which were that:

(1) Human agents can reach a justified conclusion as the result of complex deliberation.
(2) The adjustment view correctly captures (an aspect of) how humility promotes good reasoning.

We have thus identified four different ways of diagnosing the problem that the regress reveals:

(a) Explanatory failure: Human agents can reach a justified conclusion as the result of complex deliberation. However, the adjustment view fails to explain how humble reasoning can deliver a justified conclusion, so it cannot be correct. (Reject 2)
(b) Impossibility: Human agents can reach a justified conclusion as the result of complex deliberation. However, the adjustment view gives conditions for how to do so that are impossible to satisfy for a human thinker, so it cannot be correct. (Reject 2)

(c) Impossibility: The adjustment view gives correct conditions for how a humble reasoner can reach a justified conclusion. However, our assumption that it is possible for a *human* reasoner to reach a justified conclusion based on complex deliberation was incorrect. This is only possible for *ideal* reasoners. (Reject 1)
(d) Impossibility: It is impossible for any agent (ideal or nonideal) to reach a justified conclusion as the result of deliberation. We were mistaken in thinking this was possible and that such a conclusion could be reached by humble reasoning as spelled out by the adjustment view. (Rejects 1 and 2)

Depending on which of these ways of diagnosing the problem revealed by the regress we go for, different solutions present themselves. Finding the most attractive one(s) will be my task in the next section.

6. How to Get Out of the Regress

6.1 How to get out of the regress: Option (d)

Let us get option (d) out of the way first. According to (d), the regress shows that we were never aiming for a well-formulated target to begin with, because no reasoner can ever reach a justified conclusion as a result of complex reasoning. This is truly the nuclear option, as it gives up a central explanatory goal of epistemology, namely, to explain how we can have justified beliefs and credences. But besides being unattractive for that reason, it also reaches this conclusion on flimsy grounds. We considered one particular—though admittedly quite natural—way of spelling out what it means to be a humble reasoner, and we saw that it led to a regress. But we have no reason to think that the adjustment view is the only way to spell out how humility might manifest in reasoning.[11] There could be other, regress-free ways of explaining how a humble reasoner can arrive at justified conclusions.

6.2 How to get out of the regress: Option (c)

Next, consider option (c), which claims that the regress is not a problem for *ideal* thinkers and that the adjustment view correctly spells out how they should reason. To embrace this option, we need to explain two things: First, why the regress is not a problem for ideal thinkers, and second, why the view still applies to human reasoners even though they cannot comply with it.

I can think of three explanations of why the regress is unproblematic for ideal thinkers. The first claims that ideal reasoners do not need to adjust their attitudes in light of higher-order evidence that they might have made an error in reasoning, because they know their reasoning is ideal and thus error-free. Hence, the regress never even gets started. However, this line of argument is unattractive, because it makes implausible assumptions about ideal reasoners. Ideal reasoners are usually thought to be ideally rational, but not omniscient. If they were omniscient, they would not need to reason to figure things out, because they already know everything. But whether a certain piece of reasoning was flawless or not, or whether a particular agent (including oneself) is a perfect reasoner, are pieces of empirical knowledge that an ideal reasoner need not have. There is no particular empirical knowledge that ideally rational agents must have; hence, appealing to such knowledge to stop the regress is ad hoc. In other words, it is up to us how we construe ideal thinkers. We use them to illustrate what it would be like to comply with particular norms perfectly and unhindered by limitations such as time, processing capacity, or memory. We wanted to know how someone should

[11] Though see Schechter (2013) for an argument that no non-omniscient agent can ever rule out that they have misapplied an inference rule during a particular reasoning process.

reason who *cannot* be sure that their own performance is perfect. If we explain how someone should reason who *can* be sure that their own performance is perfect, we have missed our target.

The second way to argue that the regress is not a problem for ideal thinkers is to claim that they *can* actually complete it. If the regress is thought of as a supertask, that is, a task that has infinitely many sequential steps that occur sequentially in a finite amount of time, and ideal reasoners can complete supertasks, then the adjustment view could lead an ideal reasoner to a justified conclusion of their reasoning. Another possibility is that for ideal reasoners, the adjustment process can be modeled as a continuous process that converges toward a fixed point at some stage, that is, a point where the input value is mapped to itself. However, this construes the reasoning involved in a very different way from how humans would execute it. For humans, each new adjustment is a distinct reasoning process that generates distinctive higher-order evidence about its own correctness. Hence, it's not clear that such a fixed point can exist unless we change significantly how the process is modeled.[12] This general move shares with the previous one that it equips ideal thinkers with abilities that are quite different from those of human thinkers, and doing so seems solely motivated by solving the regress problem. One might argue that it is independently plausible to construe ideal reasoners as being able to do supertasks, or that their reasoning need not go through discrete steps, but a proper discussion of this would take us too far afield here. I will simply note that this solution strikes me as ad hoc, but it's not ruled out, and perhaps it's better than I give it credit here.

A third explanation of why the regress does not obtain for ideal agents strikes me as most promising. It appeals to the difference between basic and complex inferences. I briefly mentioned at the beginning of the paper that the question of how to reason humbly does not really come up for simple or basic inferences, since those are thought to confer immediate justification on their conclusions (at least in the absence of defeaters). For human thinkers, there is a small class of inferences that is thought to have this status, such as modus ponens, or-elimination, and-elimination, and a few others. What this means is that if a thinker infers something via one of these inferences, the conclusion is immediately justified. The standard contrast case is a reasoner who directly infers a complex mathematical theorem from some simple axioms. Even though this inference is valid, drawing it in one step does not result in a justified conclusion for human thinkers, and similarly for other valid, but nonobvious inferences.

The inference that is of interest for our purposes is the inference by which an agent adjusts their conclusion in light of their residual higher-order uncertainty that the conclusion might have been arrived at by bad reasoning. Is it simple or complex? Plausibly, for human agents, it is complex. It requires adjusting one's confidence in one's first-order conclusion in light of the probability that one has reasoned incorrectly, which may result in either adjusting one's confidence down, up, or not at all, depending on the case[13]. This is not a basic, one-step inference like modus ponens. However, it might be basic for ideal agents. If they do not share our cognitive limitations, the correct adjustment of their first-order attitudes might be as easy and obvious to them as a modus ponens inference is to human thinkers. But if this adjustment counts as a basic inference for ideal thinkers that confers immediate justification to the conclusion, then this inference does not generate a residual probability that the ideal agent reasoned poorly. Hence, no further adjustment is needed, and the regress does not get to the second stage.[14] By contrast, the regress continues if the reasoner is

[12]Thanks to Toby Solomon and Christian List for suggesting this.

[13]Instead of making an adjustment that outputs another precise probability, one might also suggest that adjustments should lead agents to imprecise probabilities. For example, if I initially arrived of a 0.8 credence in p, I might adjust this in light of my higher-order uncertainty about the correctness of my reasoning to an imprecise credence of [0.79, 0.81]. Further adjustments could then lead to further dilation, but the details would depend on the case and the type of error the reasoner worries they might have made. Thanks to Laura Callahan and Christian List for suggesting this.

[14]Schechter (2013) is worried that even if an inference rule is basic for an agent, the agent still cannot rule out that they have misapplied the rule. If he's right, then every inference for every agent generates residual error probabilities and the regress cannot be stopped this way. Yet, he does think that agents are entitled to "stop policing" their higher-order reasoning at some point: "A more promising picture is that we are justified in simply obeying higher-order rules that tell us how to rationally police

nonideal, because they keep making nonbasic inferences that generate residual error probabilities. This explanation of why the regress does not get past the first adjustment for ideal thinkers strikes me as less ad hoc than the first two, since having no cognitive limitations plausibly comes with an enhanced inventory of inferences that are simple or basic for the agent (see Boghossian, 2003; Dogramaci, 2012; Schechter, 2019).

Having explained how the adjustment view could apply to ideal agents, we still have to spell out how it applies to nonideal agents. Regardless of which explanation from above we favor, human reasoners can neither escape nor complete the regress; hence, they cannot arrive at a fully justified conclusion based on complex reasoning (still assuming the adjustment view is correct). However, human agents might be able to approximate the ideally justified state, which means that the adjustment view can still be thought of as a normative ideal that human agents can get close to. For example, if human reasoners can stop following the regress after a couple of steps and arrive at conclusions that are almost the same and almost as well justified as the conclusions of an ideal deliberation, then they can approximate what the adjustment view requires. Arguably, they can do so. If a human reasoner executes a complex deliberation well, and there are no misleading defeaters suggesting that they have reasoned poorly, they can typically be quite confident in having reasoned well. This means that they only need to make a small adjustment to their first-order conclusion to account for the possibility of having made an error. The subsequent adjustment to account for a possible error in that inference would be even smaller (assuming things are still going well). To be sure, this would not always be the case, but typical cases of good reasoning would plausibly proceed in this way. We could then argue that the agent is permitted to stop the regress and accept the conclusion that they have reached at that point. The explanation for this would be pragmatic: The conclusion is sufficiently good from an epistemic point of view, that is, sufficiently close to the ideal, and trying to achieve further epistemic improvements is not worth the effort. In other words, once the expected epistemic benefits of continuing the regress become too costly to be worth pursuing, the agent is permitted to stop for pragmatic reasons.[15]

A slightly different explanation of when a nonideal agent is permitted to stop the regress also appeals to the claim that in cases of good reasoning, the required adjustments get smaller and smaller. Here, the idea is that the agent may stop once "the instruments give out." The agent is permitted to stop the regress once the required adjustments become so minute that they outstrip the epistemic capacities of a human thinker.[16]

Summing up, the lesson option (c) draws from the regress is that the adjustment view can only be an account of how a humble *ideal* reasoner can arrive at a justified conclusion from complex reasoning. The most plausible explanation of why the regress is not a problem for ideal agents appeals to their expanded ability to draw simple inferences that stops the regress. Human reasoners can never be completely epistemically justified in the conclusions of their reasoning, but they can be permitted to terminate the regress for pragmatic reasons when the conclusion they reach is close enough to the ideally epistemically justified one. This option lets us accept the adjustment view, but not without costs: we must rely on assumptions about ideal reasoners that might seem somewhat ad hoc, and we have to accept a modified target of explanation. If we originally hoped to give an

---

our other rules (or perhaps still higher-order rules that tell us how to rationally police those rules or…). Such higher-order rules are not immune from potential misapplication. But we should refrain from policing them because it is too costly to do so. On this picture, the amount of rational policing that thinkers should do depends on their resource limitations. Thinkers should police as much as is reasonable, given their cognitive powers."

I agree that thinkers cannot *generally* rule out that a basic rule has been misapplied. But perhaps, in the absence of defeaters, agents are entitled to trust their application of basic rules in particular instances of reasoning.

[15]As Brian Talbot has pointed out to me, the epistemic value of getting a justified conclusion mainly comes from getting the answer to the first-order question, so the subsequent adjustments based on higher-order evidence add little to no extra epistemic value.

[16]Thanks to Charity Anderson and Peter Kung for suggesting this explanation.

account of when *human* reasoners can be fully justified in accepting the conclusions of complex deliberation, we have not reached this aim.

6.3 How to get out of the regress: Options (a) and (b)

I'll now turn to discussing options (a) and (b), which both reject the adjustment view. They thus take seriously the claim that human reasoners can be fully justified in accepting the conclusions of complex deliberation. To avoid giving up on the idea that good reasoning should exhibit the virtue of humility, we need to think more carefully about how to articulate the demands humility places on complex deliberation. How can we retain the basic insights that the adjustment view was trying to capture? We said that an agent should not accept a conclusion of a complex deliberation process unless they are sufficiently confident that they have reasoned correctly. Furthermore, if the agent has significant doubts about the quality of their first-order reasoning, they cannot just ignore these doubts and proceed with the conclusion they have reached. Simply ignoring those doubts seems dogmatic rather than humble and does not adequately take into consideration the reasoner's fallibility.[17]

The adjustment view tried to incorporate the idea that a humble reasoner should be mindful of their limitations by claiming that these limitations had to be always accounted for in the outputs of our deliberations. But this straightforwardly ignored the reasoner's inability to do so; hence, we ended up with the regress. If humility promotes good reasoning by requiring thinkers to take their limitations into account, then this rules out not just striving for certainty that one's reasoning is flawless, but it also rules out trying to complete the kind of infinite chain of inferences demanded by the adjustment view. We thus reach an interpretation of what humility demands that differs from the interpretation underlying option (c). On the current understanding, intellectual humility does not demand the impossible of *human* thinkers. That means that a human thinker may simply disregard residual error possibilities once they become insignificant enough, without this somehow compromising their humility. Humble human reasoners must reach a (possibly variable) benchmark of reasonable confidence that they have deliberated correctly, and the resulting conclusion is epistemically justified without further qualifications. The remaining small chance that they have made an error does not need to be addressed; the reasoner can ignore it without thereby lacking humility. This is captured in the

> *Benchmark view:* A reasoner may form a conclusion based on a deliberative process only if it is reasonable for them to judge that they have arrived at it through reasoning that meets a suitable benchmark of quality.[18]

This schematic definition has two important features that distinguish it from the adjustment view. The first is that it eliminates the requirement to always account for any remaining residual uncertainty about the correctness of one's reasoning. This lets it avoid the regress. The second important component is its reference to a "suitable benchmark of quality" for the agent's reasoning. It is noteworthy that I used the placeholder term "benchmark" instead of a more specific term like "confidence threshold." Part of the motivation for using this term is to ensure that the agent need not be certain that their reasoning was good, but that they still need to have good enough reasons to

[17]There are different views on what the agent should do when having significant doubts about the quality of their reasoning, but as long as the agent does not ignore those doubts, the details do not matter for the present issue. For discussion, see Staffel (2025, Ch. 9).

[18]Like the adjustment view, this is only a necessary condition for having a justified attitude based on complex reasoning, because it is silent on whether the conclusion reached by the agent is in fact supported by their total evidence or reached via a reliable process.

Furthermore, the expression "it is reasonable for them to judge" deliberately avoids requiring that the agent explicitly judges that their reasoning meets the benchmark, to avoid overintellectualizing the account.

think it was good.[19] The role of humility, if we accept the benchmark view, is then to help the reasoner set a benchmark that strikes the right balance between scrutinizing and trusting their own reasoning.

Another important reason for using the term "benchmark" is that it can be filled in different ways to make the view compatible with different types of fallibilist theories of justification in epistemology. Those theories are generally formulated as *first-order* theories of what it takes to have a justified belief (or knowledge), and they do not demand that the believer/knower must be certain. Rather, the agent is entitled to rely on their justified belief even if there is a small chance they might be mistaken. Fallibilist views have been spelled out in many different ways. One option is to introduce a (variable or fixed) confidence threshold short of certainty, such that the agent can have a justified belief or knowledge when their justified credence is above the threshold (see Dorst, 2019). Another way to do so is to articulate which error possibilities the agent needs to rule out and which ones are irrelevant or too distant to matter (Lewis, 1996). A further option is to adopt a normality condition on justified belief, according to which a belief is justified as long as it is true in the most normal worlds compatible with the agent's evidence (Smith, 2010). Yet another, more externalist option is to say that a justified belief is the product of a reliable process, where this does not require 100% reliability.

This is just a small sampling of how one might articulate a fallibilist theory of justified belief and knowledge, but all of these are *first-order* views. The benchmark view, by contrast, is a higher-order view of when the agent is permitted to stop deliberating and form a conclusion. Depending on how the benchmark view is filled, it can be given a unified treatment with a matching first-order fallibilist view of justification. For example, a threshold version of the benchmark view would say that the reasoner must be justified in having a sufficiently high level of confidence that their reasoning was correct (which may vary with context). (This is essentially a normative version of Ackerman's *Diminishing Criterion View*.) Or, a normality version of the benchmark view would say that, given the reasoner's higher-order evidence about their reasoning, they are entitled to accept the conclusion as long as it would be abnormal for their reasoning to be flawed.[20] Readers can take it from here to spell out additional versions of the benchmark view based on their favorite first-order fallibilist view of justification.[21]

How does the benchmark view relate to the image of the humble reasoner we sketched earlier? The benchmark demands that the reasoner arrive at conclusions through reasoning that meets "a suitable benchmark of quality." Intellectual humility helps us determine such a suitable benchmark—a humble reasoner sets quality standards for their own reasoning that let them address potential errors without aiming unreasonably high. By contrast, a reasoner who lacks humility might either be too easily satisfied with the quality of their reasoning, setting the benchmark too low, or they might lack self-trust, leading to excessive rumination and double-checking. Hence, being humble helps us comply with the benchmark view, which is a necessary condition for good reasoning.

However, complying with the benchmark view by itself is neither necessary nor sufficient for being intellectually humble. A humble reasoner could fail to comply with it in some cases for reasons other than a lack of virtue, for example, because they are confused or tired. Hence, perfect compliance with it is not necessary for being intellectually humble. It is also not sufficient: a

[19]Note that the definition permits, but does not require agents to form a conclusion once the benchmark is reached. It does not prevent agents from aiming higher. If we want to forbid agents from aiming higher (and possibly too high), this must be captured separately.

[20]For a more detailed discussion of how to turn first-order theories of justified belief into higher-order views along these lines, see Zhu (2025).

[21]Also, it is an open question whether this way of stopping the regress can be motivated on epistemic grounds. Depending on our explanation for how the threshold is set, it is possible that this account could collapse into the pragmatic regress stopper we formulated as part of option c, and which Schechter (2013) favors.

reasoner could always comply with the benchmark view, but exhibit a lack of intellectual humility in other ways, for example, by arrogantly overestimating their own expertise on an issue compared to that of others. This is not a problem—after all, it is implausible that meeting this narrow condition on good reasoning is all there is to intellectual humility, and also that this virtue is lost by an occasional failure to comply.

Instead, the benchmark view harmonizes with plausible views of intellectual humility, in the sense that it can capture that intellectual humility characteristically promotes good reasoning. Humble reasoners will cultivate the ability to accurately judge how well they have reasoned and when they can conclude a deliberation, and their capacity to do so will manifest itself in their compliance with the benchmark view, even if this compliance is not perfect. A lack of humility will make reasoners prone to violating the benchmark view, because they arrogantly overestimate whether their reasoning meets the benchmark for concluding. Furthermore, the benchmark view allows human reasoners to accept their limitations, because it does not require certainty in the correctness of one's reasoning, and because, unlike the adjustment view, it does not require paying excessive attention to residual error possibilities.[22]

7. Conclusion

We started with the commonsense idea that intellectual humility is a beneficial trait, because it is conducive to our epistemic aims. We then asked how this relates to being a good reasoner. Presumably, being intellectually humble should promote good reasoning, and good reasoning can be a way in which this virtue manifests itself. On the flipside, we should expect a lack of intellectual humility to be detrimental to good reasoning in some way.

We then homed in on a specific aspect of good reasoning, and asked when an agent may conclude a reasoning process. Building on some natural implications of popular accounts of intellectual humility, we developed the adjustment view. The adjustment view is supposed to capture that humble agents should not strive for an unreachable ideal of certainty that they have reasoned well, and that they should be mindful of their own limitations and fallibility. But despite being a natural way of spelling out how humility should manifest itself in complex reasoning, we saw that the adjustment view leads to a problematic regress. Hence, it fails to capture the natural idea that being intellectually humble is conducive to good reasoning, because, if the adjustment view is correct, then a human reasoner, humble or not, is unable to ever arrive at a justified conclusion of their reasoning.

We then discussed various ways of escaping the regress, including (d) rejecting the idea that any reasoners can reach justified conclusions of their reasoning, (c) claiming that the adjustment view is correct, but can only be fully complied with by ideal agents, and (a/b) rejecting the adjustment view and replacing it with an alternative that better captures the interplay between being intellectually humble and being a good reasoner. Opting for (a/b), I sketched the benchmark view, which, like the adjustment view, articulates a necessary condition for having a justified attitude based on complex reasoning. The benchmark view can be tailored to one's preferred fallibilist theory of justification, resulting in a fallibilist criterion for when an agent may accept the conclusion of a reasoning process. It successfully prevents the regress from obtaining, because the condition for being justified in forming a conclusion of a complex reasoning process can be fulfilled by human reasoners (as well as ideal reasoners), since it does not require the agent to make further adjustments to their conclusion in light of residual error probabilities. Yet, it is not entirely undemanding: it must be reasonable for agents to judge that their reasoning meets the benchmark.

[22]The benchmark view thus allows agents to trust themselves. The need for intellectual self-trust is prominently discussed and defended by Foley (2020).

The benchmark view thus meets the original explanatory target, namely to articulate when a (human) reasoner may form a conclusion of a reasoning process, and it harmonizes with prominent accounts of intellectual humility, insofar as it can explain why intellectual humility is conducive to good reasoning (and a lack of humility detrimental), and how this virtue is manifested in acts of compliance with the benchmark view.

Acknowledgments. For helpful discussion and suggestions, I would like to thank Amir Ajalloeian, Charity Anderson, Nathan Ballantyne, Bret Donnelly, Laura Frances Callahan, Will Fleisher, Javier Gonzáles de Prado Salas, Tom Kelly, Peter Kung, Maria Lasonen-Aarnio, Christin List, Juan Piñeros Glasscock, Tyler Porter, Ted Poston, Toby Solomon, Brian Talbot, Yurong Zhu, and audiences at the Arizona Conference on Epistemic Humility and Arrogance, the Orange Beach Epistemology Workshop, LMU Munich, and the Madrid Inquiry Workshop. I gratefully acknowledge funding from the John Templeton Foundation and from the Deutsche Forschungsgemeinschaft (DFG) as part of the Centre for Advanced Studies in the Humanities "Human Abilities," grant number 409272951.

Julia Staffel is a Professor of Philosophy at the University of Colorado Boulder. She is the author of two books: *Unsettled Thoughts. A Theory of Degrees of Rationality* (2019) and *Unfinished Business. Rational Attitudes in Reasoning* (2025).

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, 143(3), 1349–1368.

Ackerman, R., Binah-Pollak, A., & Lauterman, T. (2023). Metacognitive effort regulation across cultures. *Journal of Intelligence*, 11(9), 171.

Ackerman, R., & Levontin, L. (2023). Mindset effects on the regulation of thinking time in problem-solving. *Thinking & Reasoning*online first 30, 1–30.

Ackerman, R., & Thompson, V. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Topics in Cognitive Science*, 21, 607–617.

Ballantyne, N. (2023). Recent work on intellectual humility: A philosopher's perspective. *The Journal of Positive Psychology*, 18(2), 200–220.

Boghossian, P. (2003). Blind reasoning. *Supplement to the Proceedings of the Aristotelian Society*, 77(1), 225–248.

Callahan, L. F. (2024). Intellectual humility: A no-distraction account. *Philosophy and Phenomenological Research*, 108(2), 320–337.

Cameron, R. P. (2022). *Chains of Being: Infinite Regress, Circularity, and Metaphysical Explanation*. Oxford University Press.

Christensen, D. (2007). Does murphy's law apply in epistemology? *Oxford Studies in Epistemology*, 2, 3–31.

Church, I. M. (2016). The doxastic account of intellectual humility. *Logos & Episteme*, VII(4), 413–433.

DiPaolo, J. (2019). Second best epistemology: Fallibility and normativity. *Philosophical Studies*, 176, 2043–2066.

Dogramaci, S. (2012). Intuitions for inferences. *Philosophical Studies*, 165(2), 371–399.

Dorst, K. (2019). Lockeans maximize expected accuracy. *Mind*, 128(509), 175–211.

Foley, R. (2020). Self trust. In Simon, J. (ed.), *The Routledge handbook of trust and philosophy* (pp. 231–242). Routledge.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42.

Hazlett, A. (2012). Higher-order epistemic attitudes and intellectual humility. *Episteme*, 9(3), 205–223.

Hume, D. (1740/2000). In Norton, D. F. & Norton, M. J. (eds.), *A treatise of human nature*. Clarendon Press.

Huemer, M. (2016). *Approaching Infinity*. Palgrave MacMillan.

Isaacs, Y. (2019). The fallacy of Calibrationism. *Philosophy and Phenomenological Research*, 102(2), 247–260.

Law, M. K. H., Stankov, L., & Kleitman, S. (2022). I choose to opt-out of answering: Individual differences in giving up behaviour on cognitive tests. *Journal of Intelligence*, 10(86), 1–24.

Lee, D. G., Daunizeau, J. J., & Pezzulo, G. (2023). Evidence or confidence: What is really monitored during a decision? *Psychonomic Bulletin & Review*, 30, 1360–1379.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74(4), 549–567.

Meeker, K., & Poston, T. (2010). Skeptics without Borders. *American Philosophical Quarterly*, 47(3), 223–237.

Roberts, R. C., & Wood, W. J. (2003). Humility and epistemic goods. In DePaul, M. & Zagzebski, L. (eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 257–280). Oxford University Press.

Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues. An essay in regulative epistemology*. Oxford, England: Oxford University Press.

Schechter, J. (2013). Rational self-doubt and the failure of closure. *Philosophical Studies*, 163(2), 428–452.

Schechter, J. (2019). Small steps and great leaps in thought: The epistemology of basic deductive rules. In Jackson, M. B. & Jackson, B. B. (eds.), *Reasoning: New essays on theoretical and practical thinking* (pp. 152–177). Oxford University Press.

Smith, M. (2010). What Else justification could be. *Noûs*, 40(1), 139–165.

Sosa, E. (2011). *Knowing full well*. Princeton University Press.

Staffel, J. (2025). *Unfinished business. Rational attitudes in reasoning*. Oxford University Press.
Tanesini, A. (2018). Intellectual humility as attitude. *Philosophy and Phenomenological Research*, 96(2), 399–420.
Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking and Reasoning*, 20(2), 215–244.
Whitcomb, D., *et al.* (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, 94(3), 509–539.
Wieland, J. W. (2013). Infinite Regress Arguments. *Acta Analytica* 28, 95–109.
Zagzebski, L. T. (1996). *Virtues of the mind*. Cambridge University Press.
Zhu, Y. (2025). *Higher-order encroachment*. Unpublished manuscript.

Cite this article: Staffel, J. 2026. Humble Reasoning: When can I Stop?. *Canadian Journal of Philosophy* 56: e40, 1–17, doi:10.1017/can.2026.10057