DOI: 10.1111/phpr.12978

Philosophy and Phenomenological Research

ORIGINAL ARTICLE

Why bounded rationality (in epistemology)?

David Thorstad[1,2]

[1]Philosophy, Oxford University, Oxford, UK

[2]Philosophy, Vanderbilt University, Nashville, Tennessee, United States

Correspondence
David Thorstad, Philosophy, Oxford University, Oxford, UK.
Email: dthorsta@gmail.com

Abstract

Bounded rationality gets a bad rap in epistemology. It is argued that theories of bounded rationality are overly context-sensitive; conventionalist; or dependent on ordinary language (Carr, 2022; Pasnau, 2013). In this paper, I have three aims. The first is to set out and motivate an approach to bounded rationality in epistemology inspired by traditional theories of bounded rationality in cognitive science. My second aim is to show how this approach can answer recent challenges raised for theories of bounded rationality. My third aim is to clarify the role of rational ideals in bounded rationality.

1 | INTRODUCTION

Bounded rationality gets a bad rap in epistemology.[1] It is argued that theories of bounded rationality are overly context-sensitive; conventionalist; or dependent on ordinary language (Carr, 2022; Pasnau, 2013).

In this paper, I have three aims. The first is to set out and motivate an approach to bounded rationality in epistemology inspired by traditional theories of bounded rationality in cognitive science.[2] My second aim is to show how this approach can answer recent challenges raised for theories of bounded rationality. My third aim is to clarify the role of rational ideals in bounded

[1] Related issues occur in discussing the role of ideals in political theory (Barrett, forthcoming; Estlund, 2020; Simmons, 2010; Wiens, 2020). While there is no straightforward mapping between my discussion and these debates, I hope that many parts of the discussion, for example the treatment of *p*-ideals in Section 4, will find analogs in political theory.

[2] See for example Gigerenzer and Selten (2001), Simon (1955), and this paper's namesakes (Conlisk, 1996; Viale, 2021). For some recent philosophical approaches to bounded rationality see Daoust (forthcoming), Gigerenzer and Sturm (2012), Greco (forthcoming), Icard (2018); Morton (2017) Thorstad (forthcoming b) and Wheeler (2020).

rationality. I do not have the stronger aim of arguing against the need for ideal epistemology. My aims are strictly positive: I aim to set out a defensible approach to bounded rationality and make sure that this approach gets the rap sheet it deserves.

Here is the plan. Section 2 characterizes my approach to bounded rationality by way of five normative theses. Section 3 shows how this approach can respond to recent criticisms. Section 4 uses a distinction between two types of normative ideals to clarify the relationship between bounded rationality and ideal epistemology, then uses this distinction to highlight a surprising benefit of my approach: it may be our best hope for making sense of a type of full-blooded approach to Bayesian epistemology. Section 5 concludes.

2 | BOUNDED RATIONALITY

Bounded rationality is a paradigm, not a theory. It would be a category mistake to reduce bounded rationality to a list of specific normative theses. Nevertheless, we can get a good handle on the contents and motivations of my approach to bounded rationality by thinking through five characteristic normative claims made by this approach.

2.1 | Bounds matter

In practical philosophy, it is universally agreed that bounds matter to rational action. Our limited physical abilities matter: we cannot be required to manually lift a five-ton boulder, because we cannot lift it. So too, the costs of exercising our abilities matter. It might be irrational to walk from Oxford to London instead of taking the train, not because you cannot make the journey, but because the walk would consume valuable time and energy.

Bounded rationality theorists think that bounds matter to rational cognition as well. Our limited cognitive abilities matter: we cannot be required to carry out complex reasoning processes which we cannot execute. So too, the costs of cognition matter: it may be irrational to carry out complex calculations, even when these calculations are within our abilities, if those calculations take time and cognitive resources away from other inquiries and activities.[3]

Many theories of ideal rationality begin with a different starting point, which does not aim to reflect the rational importance of cognitive bounds (Carr, 2022; Smithies, 2015).[4] Theories of bounded rationality ask how agents should cognize, fixing some subset of their cognitive limitations, whereas on these views, ideal epistemology asks how agents should cognize once their cognitive limitations are removed. For example, Ralph Wedgwood suggests we picture the rational probability function by imagining "an angel perched inside the thinker's head" (Wedgwood, 2018, p. 99) who knows all facts about her mental states and carries out any needed calculations. To say that bounds matter is to hold that such theories leave out an important part of rational cognition. We are not angels but creatures with physical bodies and minds. The structure of our minds, no less than the structure of our bodies, bears on what is rational for us.

There are two ways to unpack the disagreement between bounded and ideal rationality theorists here. On a weak reading, bounded and ideal rationality are proposed as two different, and hence

[3] Which other bounds matter? I take up this question in Section 3.2.

[4] In Section 4, we will meet a different conception of ideal rationality due to Robert Pasnau (2013).

compatible types of normative assessment. Here the bounded theorist's claim is that it is important to emphasize facts about bounded rationality in order to get a full picture of human rationality.

On a strong reading, bounded and ideal rationality are not different types of assessment but rather incompatible claims about a single subject: rationality simpliciter. Just as we would not want to posit two senses of rational action, one of which abstracts away from human physical limitations and one which does not, so too on this reading the bounded rationality theorist holds that we should not posit a new type of rationality which abstracts away from human cognitive limitations. There is only one type of rationality, and it answers to cognitive bounds.

Like many bounded rationality theorists, my sympathies lie with the stronger reading, but I will leave open the weaker reading throughout my discussion.

2.2 | Procedural rationality

Herbert Simon held that a fundamental turn in the study of bounded rationality is the turn from substantive to procedural rationality (Simon, 1976). The shift from substantive to procedural rationality involves de-emphasizing normative questions about attitudes in favor of questions about the process of inquiry which produce and modify them.

Bounded rationality theorists think it is important to emphasize procedural facts because traditional norms governing doxastic attitudes either say nothing at all about cognitive bounds, or else do not make the normative impact of these bounds especially perspicuous. Consider Abelard Podgorski's discussion of rational delay: the time needed for rational agents to form attitudes on the basis of new evidence (Podgorski, 2017).[5] Many epistemological theories say nothing about rational delay: agents are required to believe what their current evidence suggests without delay. We could modify those theories to build in an acceptable interval of delay. But the most natural way to do this, Podgorski argues, would be to first ask a procedural question: how is it rational to deliberate about new evidence? Then we would ask how long this process takes and set that as the rational amount of delay. So here it looks like we must begin with procedural questions about rational deliberation before we can say anything specific about the impact of cognitive bounds such as deliberation time on rational belief, if indeed we want to allow for some such impact. And now it looks like the question about rational deliberation is in many ways prior and more revelatory in thinking about the normative impact of those bounds, since it is by asking questions about bounds on processes that we learn about the impact of those bounds on the resulting attitudes.

In epistemology, the procedural turn takes the form of a zetetic turn from a belief-focused epistemology to an inquiry-focused epistemology (Friedman, 2020). This analogy reveals two ways in which the procedural turn may be understood (Thorstad, forthcoming a).

First, we may take an *indirect* interpretation on which procedural rationality involves first assessing the rationality of processes of inquiry, then letting beliefs inherit the rational status of the processes of inquiry that produced them. Simon went in for an indirect interpretation of procedural rationality, holding that "behavior is procedurally rational when it is the outcome of appropriate deliberation" (Simon, 1976, p. 66).[6]

[5] See also Na'aman (2021a, 2021b).

[6] An indirect reading may also be a good way to unpack Podgorski's claim that norms governing states are non-fundamental. It may also be at work in Jane Friedman's suggestion that most traditional epistemological norms governing belief are incompatible with plausible zetetic norms on inquiry, because rational inquiries may produce beliefs which violate those norms (Friedman, 2019, 2020).

Second, we may take a *direct* interpretation of the procedural turn, on which attitudes and processes are separate objects of normative evaluation. On this interpretation, to ask normative questions about the rationality of cognitive processes is not yet to say anything about the rationality of attitudes that result. Then the procedural turn becomes the claim that bounded rationality should be *process-focused*, placing at least as much emphasis on normative questions about processes as on normative questions about attitudes.

In this paper, I adopt the weaker direct reading of the procedural turn, although readers sympathetic to the stronger indirect reading are welcome to read this paper with that interpretation in mind.

2.3 | Heuristic rationality

Process-focused theories of bounded rationality should say something specific about the cognitive processes that are rational for us. In particular, I claim that it is often rational for humans to cognize using a toolbox of cognitive heuristics (Gigerenzer and Selten, 2001; Gigerenzer and Gaissmaier, 2011). Most heuristics differ from non-heuristic strategies in that they process only a subset of the agent's total evidence, or draw only some of the relevant inferences warranted by that evidence.

There are at least three reasons to think that heuristic cognition is often rational. These arguments help to sort the conditions in which heuristic cognition may be rationally obligatory from those in which it is optional or impermissible.

First, there is often an *accuracy-effort tradeoff* in cognition (Johnson and Payne, 1985).[7] Investing more effort into inquiry increases the expected accuracy of our judgments, but incurs cognitive and noncognitive costs. In many circumstances, heuristics strike the best balance between accuracy and effort, performing comparably to the most demanding procedures but incurring significantly lower costs.

Second, we have *limited abilities* to perform some cognitive operations, no matter the cost. Theories of bounded rationality aim to recover a type of normative assessment in which ought implies can. In this sense, when it is beyond our ability to execute some complex nonheuristic process, then we cannot be required to do so.

Third, sometimes *less is more* (Geman et al., 1992; Gigerenzer and Brighton, 2009; Wheeler, 2020). In some circumstances, simple heuristic rules outperform more complex nonheuristic rules by avoiding overfitting. In these situations, heuristics can be rational even if costs and abilities are not at issue.

Why is it important to study heuristic cognition? In Section 2.5, we will meet one surprising reason for epistemologists to study heuristics: the study of rational heuristic cognition may put pressure on the truth or importance of traditional epistemological norms.

2.4 | Ecological rationality

Herbert Simon held that human rationality is shaped by a pair of scissors, whose blades are the agent's internal cognitive limitations and the structure of her environment (Simon, 1990). In thinking about rationality it is easy to focus on the first blade and ignore the second. Theories

[7] There is often, but not always an accuracy-effort tradeoff. That is the lesson of less is more effects.

of *ecological rationality* insist, by contrast, that rationality is ecological or environment-relative (Douven, 2020; Todd and Gigerenzer, 2012). Humans are bounded by our environments as much as by our internal cognitive structure. A good theory of bounded rationality should incorporate both bounds.

Broadly speaking, there are two reasons to take an ecological approach to bounded rationality. First, as we learned from reliabilists, it is misleading to assess strategies by looking at their performance in a single instance (Goldman, 1986). That is largely a matter of luck, and says little about the strategy itself. To get a luck-free assessment of cognitive processes, we need to look at their performance across an environment of similar problems.

Second, all of the best-known heuristics work well in some environments and poorly in others. For this reason, it makes little sense to ask whether a given heuristic is rational or irrational full-stop. Rather, we must ask: in which environments would this heuristic be a rational strategy? That is the question which ecological rationality poses.

Ecological rationality is instrumental in resisting empirically-driven challenges to the rationality of heuristic cognition. For any given heuristic, it is easy to find conditions under which that heuristic will perform poorly. Indeed, a full theory of ecological rationality will precisely describe those conditions. If that is right, then to show that heuristics sometimes perform poorly under laboratory conditions is not yet to question the rationality of those heuristics. It must be shown that the heuristics frequently break down under the conditions where they are proposed for use. This takes us to our final claim.

2.5 | Vindicatory epistemology

A bevy of laboratory experiments reveal that agents sometimes make judgments which violate traditional rationality requirements (Fischoff and Broomell, 2020; Gilovich and Griffin, 2002; Kahneman et al., 1982). It is natural to treat these findings as instances of irrational cognition, and sometimes that may be the right reaction.

The program of *vindicatory epistemology* (Thorstad, forthcoming c) aims to show that many violations of traditional rationality requirements occur as a result of boundedly rational deliberation.[8] We violate requirements such as probabilistic coherence or deductive closure because we are deliberating in the most rational way possible given our bounds, and the procedures that are rational for us to use lead occasionally to violations of traditional norms.

When this is the case, an indirect interpretation of the procedural turn implies that beliefs which violate traditional rationality requirements can be rational. A direct interpretation may grant that the attitudes in question are irrational, but holds that this assessment is incomplete and misleading. What should be emphasized in such cases is that the attitudes in question, whatever their rational status, resulted from fully rational heuristic inquiry, and could only have been avoided by adopting wastefully irrational nonheuristic methods. In this way, an indirect interpretation of vindicatory epistemology suggests that traditional epistemological norms may be false, while a direct interpretation suggests that they are importantly incomplete (Thorstad, forthcoming a).

In this section, I have set out and motivated the approach to bounded rationality that I defend by way of five characteristic normative theses. Bounds matter to rational cognition. It is

[8] For recent philosophical exemplars see Hedden (2012); Morton (2017); Polonioli (2013) and Icard (2018). In the recent empirical literature, see for example Lieder et al. (2018); Schooler and Hertwig (2005) and Vul et al. (2014).

important to take a procedural lens towards rationality, and the procedures that are rational for us are often heuristic. Rational cognition responds to environmental bounds as well as internal cognitive bounds. And the right approach to bounded rationality can vindicate many seeming irrationalities as the results of boundedly rational deliberation. In the next section, I show how my approach avoids two problems recently raised for theories of bounded rationality.

3 | CONVENTIONALISM AND CONTEXT-SENSITIVITY

A recent paper by Jennifer Carr argues that theories of bounded rationality struggle to generate a type of normative assessment that is neither conventional nor seriously context-sensitive (Carr, 2022). In this section, I argue that my view avoids both challenges.[9]

3.1 | Conventionalism

For Carr, *conventions* are "regularities in behavior that serve some coordinative function, where some alternative regularity in behavior would have served the same coordinative function equally well, if widely adopted" (Carr, 2022). Carr considers a conventionalist metaepistemological view: Sinan Dogramaci's epistemic communism (Dogramaci, 2012, 2015, 2017). On this view, societies use epistemic assessments to coordinate on a fixed stock of belief-forming rules for promoting the efficient formation of true beliefs about matters of interest via testimony. The function of normative evaluations is to promote compliance with accepted belief-forming rules, censure non-compliant agents, and identify compliant agents whose testimony can be trusted to flow from reliable belief-forming rules.

Carr identifies three plausible sources of conventionality in the rules sanctioned by epistemic communism. First, there are many equally truth-conducive sets of belief-forming rules, so it will be an arbitrary matter which of these sets a given society coordinates on. Second, bounded agents need to winnow large sets of belief-forming rules down to a size and content compatible with their epistemic limitations. Plausibly, there are multiple incompatible ways to perform this winnowing. Third, coordinating rules across groups of agents may require reducing the number of sanctioned rules in order to keep epistemic assessments manageable. Again, there may be multiple equally-good ways to perform this winnowing. Carr uses this discussion to suggest that theories of bounded rationality will be unacceptably conventionalist.

I am inclined to grant that Dogramaci's communist picture of epistemic evaluation leads to a strongly conventionalist picture of epistemic normativity. I suspect this will come as little surprise to Dogramaci, who holds precisely on these grounds that there can be no general theory of epistemic rationality (Dogramaci, 2015). Dogramaci and other conventionalists may reply to Carr by

[9] In what ways does my considered view differ from Carr's notion of ideal rationality? I think it probably differs in at least the following ways: it makes cognitive limitations normatively relevant; accepts that ought implies can; and is a form of bounded rationality and satisficing. My view may also differ from Carr's notion of ideal rationality in some or all of the following ways: it is more empirical and ameliorative; treats heuristics as more rational and reliable; does not involve superbabies; is less friendly to decision theory; and treats normative ideals as normative standards rather than p-ideals (see Section 4). However, I am not sure if we differ in these respects, and in any case my aim is not to emphasize disagreements with Carr but rather to characterize, motivate and defend an approach to bounded rationality. Agreement on some aspects of rationality is to be celebrated, not spurned.

defending conventionalism. But for the purposes of this paper, I want to stress a different line of response: there is no direct link between bounded rationality and conventionalism.

The account of bounded rationality in Section 2 contains not a whiff of conventionalism. We could make it conventionalist by coupling it with a theory such as epistemic communism. But we could also make it non-conventionalist by taking an accuracy-first approach on which agents should cognize so as to best promote accuracy among their own beliefs (Karlan, 2021); a coherence-based approach on which they should be as coherent as possible (Staffel and De Bona, 2018); or a consequentialist approach on which they should cognize in the most value-promoting ways (Thorstad, forthcoming c). Now readers may think that epistemic communism is the correct metaepistemological theory, in which case they are free to adopt conventionalist versions of any of the above principles. But it would be patently unfair to turn around and blame conventionalism on the bounded rationality theorist, because many bounded rationality theorists are not conventionalists.

In fact, there are at least two reasons why bounded rationality theorists may resist conventionalism. I do not take either of these considerations to show that bounded rationality theorists cannot be conventionalists.[10] But together, they will put pressure on the association between bounded rationality and conventionalism.

First, Carr reads Dogramaci as holding that communal rules should make no allowance for individual variation in ability. If some individual can apply more demanding heuristic or non-heuristic methods than others or can apply these methods at lower cost, this fact does not make those methods any more rational for her. I am not sure if this is the right reading of Dogramaci. But if it is, this implication is at odds with leading approaches to bounded rationality, which take pains to stress that what is rational for an agent depends on her abilities and the costs of exercising them. Just as having more money may permit us to buy champagne instead of water and having more stamina may permit us to run instead of walking, having the capacity to think longer and harder before forming a judgment can make it rational for us to think longer and harder.

Second, heuristic cognition is often difficult to introspect.[11] It is not always possible for agents themselves, let alone external observers, to identify the heuristic processes used to produce judgments. And when it is possible to identify heuristics, this often requires a detailed laboratory study. Because identifying bounded agents' cognitive processes is often impossible or expensive, efficient coordinative conventions may not always be able to take detailed account of the nuances of agents' cognitive processes in determining whom to praise, and whom to rebuke. This consequence would be at odds with standard interpretations of procedural rationality, on which understanding the precise structure of agents' cognitive processes is crucial to assessing their rationality. And it may cause trouble for vindicatory epistemology, which uses highly detailed examinations of cognitive processes to challenge allegations of irrational cognition.

Again, my purpose in this section is not to challenge epistemic conventionalism. My aim is to break the association between conventionalism and bounded rationality. Many approaches to bounded rationality, including my own, are not conventionalist. And there are some reasons for bounded rationality theorists to resist conventionalism. Some readers may nonetheless be attracted to conventionalism. But conventionalism is not a necessary feature of bounded rationality.

[10] Most prominently, perhaps Stich (1990) might be read as a conventionalist, insofar as he allows that different societies may adopt different normative assessments depending on their values and their interpretation of the truth predicate.

[11] I would be remiss if I did not add that paradigmatic heuristics can be executed consciously. But it is usually not efficient to do so.

3.2 | Context-sensitivity

For Carr, a type of evaluation is *seriously context-sensitive* if "there is no normatively privileged resolution of one or more of the context-sensitive parameters for evaluations of that kind" (Carr, 2022). Carr thinks that bounded rationality is seriously context-sensitive because she holds that there is no normatively privileged way of distinguishing the bounds that are held fixed in normative assessment from those that are not.

To illustrate, Carr considers the following lists of bounds:

> *Cognitive limitations that lower the bar for nonideal rationality*: our limited computational power, informational storage, processing speeds, integration of different cognitive systems, information retention ...
>
> *Cognitive limitations that don't*: our dispositions towards implicit biases, unreliable heuristics, delusional reasoning, misinterpreting statistical phenomena as having causal explanations, inflated sense of one's own driving ability, overoptimism/pessimism, overestimating the moral superiority of one's own side in a fight with a spouse, family member, or departmental faction. (Carr, 2022).

Carr considers and rejects two proposals for what could ground a normative distinction between the former and latter bounds. Finding both proposals wanting, Carr concludes that the line between the two categories is blurry: "whatever sharp line nonideal epistemologists draw, it could have been drawn elsewhere" (Carr, 2022).

Now I am not sure that I can provide a perfectly bright line between bounds that matter and bounds that do not. But we do not usually demand a bright line between two categories to accept that there is a well-formed distinction between them. For example, most philosophers would accept that there is a distinction between culpable and nonculpable ignorance, but many are unsure where that line is to be drawn. Likewise, I do not aim to quell disagreement about which bounds matter. The distinction between bounds that matter and bounds that do not, like the distinction between culpable and nonculpable ignorance, is a matter of substantial controversy. But Carr is within her rights to demand a well-motivated proposal for what could distinguish the bounds on her list. That would put the burden back on opponents to produce new grounds for taking the distinction between bounds that matter and bounds that don't matter to have no normatively privileged resolution. Where could we find such a proposal?

We saw in Section 2 that bounded rationality theorists are concerned with two sorts of bounds: internal bounds imposed by an agent's internal cognitive structure, and external bounds imposed by her environment. The bounds on Carr's list are all internal bounds, so my aim will be to give an account of which features of an agent's internal cognitive structure matter normatively.

One of the most fundamental distinctions in cognitive science is the distinction between an agent's fixed cognitive architecture and the representations or processes realized within that architecture. Here is how a recent review draws that distinction:

> A cognitive architecture specifies the underlying infrastructure for an intelligent system. Briefly, an architecture includes those aspects of a cognitive agent that are

> constant over time and across different application domains. (Langley et al., 2009, p. 141).

For example, it is an architectural fact that our working memory has a fixed capacity, but a non-architectural fact which beliefs are currently held in working memory. My proposal is that the features of an agent's internal cognitive structure which must be held fixed during normative assessment are the totality of facts about her token cognitive architecture.[12] This proposal has three advantages.

First, it is well-motivated. It is very natural to interpret talk of an agent's internal cognitive structure as talk about cognitive architecture. We will see in Section 4 that this is how many Bayesians have interpreted internal bounds. This proposal also captures the view that ought implies can, which Carr takes to be one of the fundamental motivations for bounded rationality. Because we cannot, at the moment of choice, change our cognitive architecture, the question facing bounded agents is what we ought to do given the architecture that we have, not what we might prefer to do if a different architecture were available to us. And there is a striking analogy between this proposal and the case of rational action, in which it is standardly thought that an agent's physical architecture, such as her size and strength, should be held fixed in normative evaluation. To determine whether I should lift a weight, we must hold fixed facts about physical architecture such as my having two arms and the motions and exertions required to lift a weight with my arms. Similarly, to determine whether I ought to implement some cognitive process, we must hold fixed whether I can implement it and what would be required to do so, given my cognitive architecture.

A second advantage of my proposal is that it is well-understood. Because the distinction between architectural and non-architectural features is one of the fundamental tools of modern cognitive science, it should take a great deal to convince us that there is no privileged resolution of this distinction. That is not to say that there is no disagreement about where the line should be drawn, but it should take a great deal of evidence to convince us that there is anything deeper than ordinary scientific disagreement at play here.

A final advantage of my proposal is that it correctly separates the items on Carr's list. Begin with the cognitive limitations that are held fixed in normative evaluation: limited computational power, informational storage, processing speeds, integration of different cognitive systems, and information retention. The first four of these limitations are paradigmatic examples of architectural features and should therefore be fixed. The last, information retention, must be spelled out with some care. If it picks out broad architectural facts such as capacity limits on working memory, then my view correctly holds information retention fixed. But if information retention refers to specific facts within a cognitive architecture, for example the fact that I forgot my partner's birthday, that is not an architectural fact. Some philosophers may think that facts about memory are inapt for rational assessment, and my view does not settle that question. But if facts about memory are apt for rational assessment, my view correctly puts agents on the hook for items of information that they have forgotten.

[12] For this to be plausible, we must distinguish between broad architecture *types*, such as ACT-R (Anderson, 1990), and specific architecture *tokens*, such as a specific Lisp instantiation of ACT-R for a given application. Token architectures must be specified in enough detail to relevant bounds such as cognitive costs and processing power that are relatively constant across time and application domains. Here we might look for inspiration to Bayesian understandings of cognitive architecture (Anderson, 1990; Howes et al., 2009; Lieder and Griffiths, 2020), and in particular to the analysis of Howes et al. (2009) which also proposes to identify normative relevant internal bounds with the bounds imposed by cognitive architecture. It might be fruitful for further work to explore the notion of token architectures in more detail. Thanks to an anonymous referee for pushing me to clarify my views here.

Now turn to the bounds not held fixed. With the exception of implicit bias, these bounds fall into two categories. The first category contains specific processes which can be implemented within a cognitive architecture: unreliable heuristics and delusional reasoning. Because our architecture makes it possible for us to implement other strategies such as reliable heuristics and non-delusional reasoning, my view correctly puts agents on the hook for poor strategy choices.[13] The second category contains attitudes which result from our processing choices, such as overoptimism and an inflated sense of one's own driving ability. Our architecture makes it possible to reason towards different attitudes on the basis of evidence, and for that reason these attitudes should not be held fixed, but treated as normatively-assessable choices within a fixed cognitive architecture.

The case of implicit bias is more complicated. On the one hand, although implicit biases may well be unconscious, it is rarely suggested that implicit biases are features of cognitive architecture, and for that reason my view is friendly towards the idea that implicit biases may be rationally assessable. On the other hand, the rational status of implicit biases turns on many questions that are orthogonal to our discussion. For one thing, there is increasing disagreement about whether current conceptions of implicit bias capture the phenomenon that we are after (Mandelbaum, 2016; Holroyd, 2016; Oswald et al., 2013). For another, many discussions of implicit bias argue that implicit biases are culpable or blameworthy, but not that they are irrational (Holyrod et al., 2017), and some deny that all implicit biases are even blameworthy (Saul, 2013). Many moves are possible here, and there is good reason to think that some of these moves will require small revisions to many traditional normative theories, including theories of bounded rationality. But we have not yet been given any reason to think that bounded rationality theorists are in any worse position to accommodate these moves than any others. Indeed, bounded rationality theorists may be in a better position here than most, for we do not insist that attitudes or processes need to be conscious or introspectable in order to be rationally assessable.

Now I am sure that some readers may disagree with my claim that bounds matter just in case they are features of cognitive architecture. I think that there is much to be said in favor of that claim: it is well-motivated, well-understood, and divides Carr's cases correctly. But my aim is not to end discussion of which bounds matter normatively. Quite the opposite, my aim is to suggest that there may well be a normatively privileged way to settle the issue and to invite further discussion aimed at settling it.

In this section, we have seen how bounded rationality theorists can meet the charges of conventionalism and context-sensitivity. In the next section, I reflect on the role of rational ideals in bounded rationality, and use this distinction to highlight a surprising connection between bounded rationality and Bayesian epistemology.

4 | BOUNDED RATIONALITY AND IDEAL RATIONALITY

What is the relationship between ideal rationality and bounded rationality? On many accounts such as Carr's, the distinction between bounded and unbounded rationality coincides with the distinction between non-ideal and ideal rationality. For this reason, I have passed freely between the two terms in Sections 1-3.

[13] We learn to select good strategies through metacognitive strategy selection (Lieder and Griffiths, 2017; Marewski et al., 2010).

In this section, I consider a different account of ideal rationality due to Robert Pasnau (2013) on which ideal rationality absorbs some lessons from the bounded approach. I distinguish two notions of a cognitive ideal in order to contrast the role of ideals in bounded rationality and Pasnau's ideal rationality (§4.1). Then I argue that the bounded approach has at least one surprising advantage: contrary to appearances, my account of bounded rationality is better-suited than Pasnau's ideal rationality to make space for full-blooded versions of Bayesian epistemology (§§4.2-4.3).

4.1 | Two types of ideals

For Pasnau, ideal epistemology proceeds in two stages. The first stage, *idealization*, specifies "what would count as perfection for beings such as us, in a world such as ours" (Pasnau, 2013, pp. 1005-6). Unlike Carr's approach, Pasnau's idealization stage does not abstract away from cognitive or environmental bounds: it aims to specify the most perfect state we can achieve given our bounds. Call this a *p-ideal*.

By way of example, Pasnau suggests that we read the Cartesian notion of *scientia* as describing a p-ideal of certain and evident cognitions, built upon a foundation of certain and evident cognitions, all of which can be grasped by the knower. Here the suggestion is not that rationality always requires us to achieve *scientia*, nor even that *scientia* is achievable in every situation, but only that *scientia* is the most perfect state we can achieve.

The second stage, *application*, considers "how much of the human epistemic ideal might reasonably be applied to ordinary cognitive agents in everyday life" (Pasnau, 2013, p. 1006). For example, consider the debate between internalism and externalism about knowledge. Externalism, Pasnau suggests, lies further than internalism from the ideal of *scientia*, but does it lie too far away to count as knowledge? Pasnau suggests we can get a grip on this question by asking whether it is ideally possible to establish the reliability of the senses on a non-circular internalist footing. If so, then that may be some evidence in favor of an internalist conception of knowledge, but if not we had better make space for externalism on pain of skeptical paralysis.

Now bounded rationality is a thesis about rationality, not knowledge. Bounded rationality takes no stance on the role of p-ideals in the analysis of knowledge. But in assessing rationality, there is a viable alternative view. On this view, theorizing begins by specifying a *normative standard* which directly pronounces on the rationality of token attitudes or processes. For example, we may say that rationality requires us to promote as much accuracy (Ahlstrom-Vij and Dunn, 2018), coherence (Staffel, 2020) or knowledge as possible; to honor rather than promote these goals (Sylvan, 2020); or to best satisfy a collection of structural rationality conditions (Broome, 2013). The second stage, application, begins by precisely specifying the normatively relevant features of an agent's cognitive situation, such as her environment and cognitive architecture, then asking which processes or attitudes are permissible in this situation by the light of the relevant normative standard. For example, an accuracy-first approach to rational inquiry might require agents to use the most reliable inference rule at their disposal.

On this construal, ideal rationality begins by specifying the most perfect state attainable by humans, a p-ideal, and then asking how much of that ideal can be required in different situations. By contrast, bounded rationality begins with a normative standard, which does not describe a state of an agent but rather directly says how the rationality of token attitudes or processes is to be judged.

Does bounded rationality deny the need for rational ideals? Bounded rationality does not appeal to p-ideals, but it may be productive to regard normative standards as another type of rational ideal. The question remaining is: why think that bounded rationality is determined by normative standards rather than p-ideals? While a full answer to this question is beyond the scope of the present discussion, in the rest of this section I suggest one reason to prefer normative standards over p-ideals: it is our best hope for a defensible, full-blooded approach to Bayesian epistemology for bounded agents.

4.2 | Full-blooded Bayesianism

What do we do when we do Bayesian epistemology? Pasnau suggests that we cannot take formal models of cognition as literal descriptions of how bounded agents do or should cognize:

> Bayesianism, and other such technical approaches ... describe a methodology that ordinary agents could not possibly pursue, because of the overwhelming mathematical complexity that would attend to its application to any real world case. (Pasnau, 2013, p. 1101).

Pasnau suggests instead that we construe formal epistemology as describing p-ideals, and consider what applications these ideals might have for ordinary agents.

Consider the case of Bayesian epistemology. Certainly many epistemologists are not Bayesians, and some Bayesians would endorse Pasnau's injunction against taking Bayesian models literally. But a striking fact is that many Bayesian epistemologists have taken their models to describe normative standards, such that agents who fail in any way to meet these standards are cognizing irrationally (Hartmann and Sprenger, 2010; Joyce, 1998; Staffel, 2020). And as a descriptive matter, Bayesian modeling is among the fastest-growing areas of cognitive science. While it is an open debate how descriptive Bayesian models are to be taken (Bowers and Davis, 2012; Colombo et al., 2020), many Bayesians have stood their ground and taken Bayesian models to provide good descriptions of how agents like us actually cognize (Clark, 2013; Howhy, 2013; Rescorla, 2019).

Let *full-blooded* Bayesianism be the view that Bayesian norms describe how agents like us ought to cognize, and often provide good descriptions of how we in fact cognize. Full-blooded Bayesianism is a nonstarter if we take Bayesian norms to express p-ideals rather than normative standards. It is often thought that full-blooded Bayesianism is also incompatible with mainstream theories of bounded rationality. In this section, I hope to show that the opposite is true: while bounded rationality theorists need not be Bayesians, the most defensible version of full-blooded Bayesianism makes detailed appeal to bounded rationality.[14]

[14] Where, if at all, do bounded rationality theorists disagree with full-blooded Bayesians? We certainly don't disagree on the need for normative standards, and in principle, bounded rationality theorists can accept many normative standards such as accuracy- or value-promotion, although we are usually hesitant to accept coherence-based standards (Arkes et al., 2016). Bayesians following the vindicatory strategies in this section may account for much of bounded rationality through appeal to sampling (Icard, 2018), heuristics (Oaksford and Chater, 2007), incomplete updating (Dallman, 2017), limited attention (Sims, 2003) and other concessions to bounded rationality. There will remain debates about the explanatory status of Bayesian theorizing (Bowers and Davis, 2012; Jones and Love, 2011; Colombo and Hartmann, 2017) and the ability of Bayesians to cover all cases, such as metacognition (Proust, 2013). But many theorists hold out hope for a fully successful Bayesian theory of bounded rationality, and that hope is something to celebrate rather than sabotage.

To see why full-blooded Bayesians increasingly turn towards bounded rationality, consider the method of *rational analysis* (Anderson, 1990; Chater and Oaksford, 1999) that underwrites modern approaches to Bayesian cognitive science. This approach begins by specifying normative standards for cognition, in the form of goals to be promoted, then making the defeasible supposition that agents cognize as they ought, testing and modifying models as necessary. Here is a paradigmatic six-step statement of the method of rational analysis (Chater and Oaksford, 1999, p. 59):

(1) *Goals*: specify precisely the goals of the cognitive system.
(2) *Environment*: develop a formal model of the environment to which the system is adapted.
(3) *Computational limitations*: make minimal assumptions about computational limitations.
(4) *Optimization*: derive the optimal behavior function, given 1-3 above.
(5) *Data*: examine the empirical evidence to see whether the predictions of the behavior function are confirmed.
(6) *Iteration*: repeat, iteratively refining the theory.

There is a lot to like in this view. For example, Step 2 incorporates the claim that rationality is ecological. But Bayesian rational analysis has been widely criticized for its third step, which asks us to make minimal assumptions about computational limitations. Why not instead make accurate assumptions about agents' computational limitations?

Rational analysis quickly granted this criticism. They accepted that, as a normative matter, agents should respond to the bounds they actually have, and as a descriptive matter, that is what agents often do. The paradigm of *cognitively bounded rational analysis* (Howes et al., 2009) replaces the third step of rational analysis with an explicit insistence on modeling cognitive bounds. That emphasis has been shared by most recent Bayesian paradigms in cognitive science.[15]

These paradigms often agree not only with the spirit of my bounded approach, but also with its letter. Howes and colleagues hold that a system exhibits cognitively bounded rationality when "its behavior maximizes subjective expected utility given the constraints on the cognitive architecture and the local task environment" (Howes et al., 2009, p. 728). That is precisely the account I gave in Section 3.2 of which bounds matter, coupled with a specific normative standard, subjective expected utility maximization, by which the performance of bounded agents is to be assessed. This account could be generalized by coupling it with other normative standards, such as expected- or actual-accuracy maximization.

By building bounds directly into normative models, we create the possibility of a full-blooded Bayesian approach on which agents do and should comply with Bayesian norms. But what of the criticism that Bayesian norms themselves are too demanding for bounded agents to execute? Here, again, full-blooded Bayesians have thought that their best strategy is to incorporate bounds directly into Bayesian normative standards.

[15] These include computational rationality (Gershman et al., 2015), boundedly rational analysis (Icard, 2018), and resource-rational analysis (Lieder and Griffiths, 2020).

4.3 | An example: Sampling and matching bias

Consider, by way of illustration, one way in which Bayesian norms look too demanding: they ask agents to make even the smallest judgments on the basis of their total stock of relevant evidence. But it is hard to see how bounded agents could afford to call up and analyze more than a small subset of that information on a regular basis. How can Bayesians incorporate the need to make judgments based on a representative subset of evidence?

An increasingly popular thought among Bayesians is that rational agents draw *samples* of information from memory rather than surveying all of that information at once (Icard, 2018; Stewart et al., 2006; Zhu et al., 2020). Properties of the samples, such as their mean and variance, are taken as proxies for properties of the underlying distribution in the usual way. For example, asked to categorize a distant tree you might draw samples from your underlying beliefs about its possible composition. The category most frequently represented in drawn samples would be returned as the category of the tree. For example, if 70% of the drawn samples were pine trees, then you would judge that the tree is a pine.

The point is that in many contexts, accurate judgments can be made on the basis of a small number of samples. When that is right, there is no great mystery as to how bounded agents could access enough information to make Bayesian calculations. After all, most non-Bayesian models of cognition also require agents to draw samples of information from memory before making a judgment. Nor is there any great mystery about how the computations described could be carried out. Most of us can categorize items by counting the number of times each category appears in a sample.

This type of bounded turn not only improves the prospects for full-blooded Bayesians, but also suggests vindicatory verdicts about rationality that will be difficult to recover by reflecting on p-ideals. By way of illustration, one of the most persistent findings across many domains of cognition is a *matching bias* in which judgments are made in proportion to probabilities (Vulkan, 2000). Asked to categorize a tree on the basis of evidence making it 70% likely that the tree is a pine and 30% likely that the tree is a birch, agents exhibiting matching bias will become 70% likely to categorize the tree as a pine and 30% likely to call it a birch. That is not what traditional normative theories demand. If forced to categorize this tree, agents should always categorize it as a pine, not merely become more likely to do so as the probability that it is a pine increases.

Classically, matching bias might be interpreted as a sign of irrationality, but that judgment may be too hasty. Suppose it is rational for agents to make a given categorization judgment by sampling rather than surveying entire probability distributions. If judgments are made on the basis of a single sample, we recover a matching bias: 70% of samples drawn will be pines, so agents making judgments by sampling will exhibit complete matching bias. When cognitive costs are high enough or time is tight, it may be rational to make judgments on the basis of a single sample (Icard, 2018; Vul et al., 2014), in which case we get the result that matching bias can result from fully rational inquiry. More generally, building the process of sampling into normative models together with bounds such as computational costs allows us to say in detail how many samples of information it will be rational to draw in any given situation. As more samples are drawn, we expect a small but rapidly diminishing amount of matching bias to remain, allowing us to say in great detail how much matching bias should be expected to result from rational inquiry in a given case.

This is exactly the sort of vindicatory prediction that the bounded rationality theorist promised us, in which a bias that appears to be the result of sloppy and irrational inquiry is seen to result

from rational inquiry once relevant bounds are included into normative models. Far from a criticism of full-blooded Bayesianism, this model was developed by mainstream Bayesian epistemologists and cognitive scientists. This prediction relies essentially on considerations of bounded rationality. And it is not a prediction that can easily be derived by reflection on p-ideals. For example, since humans are capable of surveying most or all information held in memory, we might take the p-ideal in this case to be one in which most information held in memory is sampled. But then one-shot sampling lies as far as possible from the relevant p-ideal. More generally, it is hard to see how we could explain the rationality of one-shot sampling by reflection on p-ideals alone, without passing through a story about normative standards involving evidence, costs and bounds. And once we do that, it becomes tempting to take p-ideals out of the story and just directly ask what the relevant normative standards require in any given case.

In this section, we distinguished between two conceptions of normative ideals: p-ideals and normative standards. We saw that bounded rationality does not make use of p-ideals, but does make use of normative standards. We saw how coupling bounded rationality to a reliance on normative standards rather than p-ideals makes room for a surprising alliance between bounded rationality and full-blooded Bayesian epistemology. And we saw how this alliance can be used to begin the vindicatory project of rehabilitating seeming irrationalities as the result of boundedly rational deliberation.

5 | CONCLUSION

In this paper, I have done three things. First, I set out and motivated an approach to bounded rationality by way of five characteristic normative theses. Second, I showed how that approach can respond to the criticisms that bounded rationality is overly conventionalist and context-sensitive. Third, I distinguished between two conceptions of normative ideals and used that distinction to motivate an alliance between bounded rationality and full-blooded Bayesian epistemology. I suggested that the best hope for full-blooded Bayesians is to turn towards bounded rationality, and to treat their view as characterizing normative standards rather than p-ideals.

This discussion leaves open important questions for future research. One question concerns the extent of vindicatory epistemology: just how much of human cognition can be rationalized (Cohen, 1981; Lieder and Griffiths, 2020)? Relatedly, how should vindicatory epistemology be interpreted? Should we take a direct interpretation on which it reveals traditional epistemic norms to tell an incomplete story about rational cognition, or an indirect interpretation on which vindicatory epistemology falsifies traditional epistemic norms? The choice between direct and indirect interpretations is also relevant to the epistemology of inquiry: here, too, considerations of bounded rationality have been used to suggest that rational inquirers may sometimes violate traditional doxastic norms (Friedman, 2020). Does this show traditional doxastic norms are false or rather that they tell an incomplete story about human rationality (Thorstad, 2021, forthcoming b)?

It may also be productive to explore stronger defenses of bounded rationality. In this paper, I have left open an interpretation on which bounded rationality and ideal rationality are complementary forms of normative assessment. But on a stronger reading, bounded rationality and ideal rationality are inconsistent theories about the same object: rationality. On this reading, defenses of bounded rationality are meant to show that ideal rationality is false or unnecessary. What, if anything, could be said in favor of this stronger project?

As we approach these questions, it is important to be clear about what is meant by bounded rationality and ideal rationality in order to make sure that both approaches get the rap sheet that they deserve.[16]

ORCID

David Thorstad https://orcid.org/0000-0003-1985-8931

REFERENCES

Ahlstrom-Vij, K. and Dunn, J., editors (2018). *Epistemic consequentialism*. Oxford University Press.

Anderson, J. (1990). *The adaptive character of thought*. Psychology Press.

Arkes, H., Gigerenzer, G., and Hertwig, R. (2016). How bad is incoherence? *Decision*, 3(1):20–39.

Barrett, J. (forthcoming). Deviating from the ideal. *Philosophy and Phenomenological Research*, page forthcoming.

Bowers, J. and Davis, C. (2012). Is that what bayesians believe? reply to griffiths, chater, norris and pouget (2012). *Psychological Bulletin*, 138(3):423–426.

Broome, J. (2013). *Rationality through reasoning*. Wiley.

Carr, J. (2022). Why ideal epistemology? *Mind*, 131(524):1131–62.

Chater, N. and Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3(2):57–65.

Clark, A. (2013). Wherever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36:181–204.

Cohen, J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4(3):317–331.

Colombo, M., Elkin, L., and Hartmann, S. (2020). Being realist about bayes, and the predictive processing theory of mind. *British Journal for the Philosophy of Science*, 72(1):185–220.

Colombo, M. and Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *British Journal for the Philosophy of Science*, 68:451–84.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, 34:669–700.

Dallman, J. (2017). When obstinacy is a better (cognitive) policy. *Philosophers' Imprint*, 17(24):1–18.

Daoust, M.-K. (forthcoming). The comparison problem for approximating epistemic ideals. *Ratio*, page forthcoming.

Dogramaci, S. (2012). Reverse engineering epistemic evaluations. *Philosophy and Phenomenological Research*, 84(3):513–530.

Dogramaci, S. (2015). Communist conventions for deductive reasoning. *Noûs*, 49:776–799.

Dogramaci, S. (2017). Why is a valid inference a good inference? *Philosophy and Phenomenological Research*, 94(1):61–96.

Douven, I. (2020). The ecological rationality of explanatory reasoning. *Studies in History and Philosophy of Science Part A*, 79:1–14.

Estlund, D. (2020). *Utopophobia: On the limits (if any) of political philosophy*. Princeton University Press.

Fischoff, B. and Broomell, S. (2020). Judgment and decision making. *Annual Review of Psychology*, 71:331–355.

Friedman, J. (2019). Checking again. *Philosophical Issues*, 29(1):84–96.

Friedman, J. (2020). The epistemic and the zetetic. *Philosophical Review*, 129(4):501–36.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58.

Gershman, S., Horvitz, E., and Tenenbaum, J. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–8.

Gigerenzer, G. and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1):107–43.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62:451–82.

[16] Thanks to Marc-Kevin Daoust, Tomi Francis, Natasha Oughton and two anonymous referees for this journal for their feedback on this paper.

Gigerenzer, G. and Selten, R., editors (2001). *Bounded rationality: The adaptive toolbox*. MIT press.
Gigerenzer, G. and Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese*, 187:243–68.
Gilovich, T. and Griffin, D. (2002). Heuristics and biases: Then and now. In Gilovich, T., Griffin, D., and Kahneman, D., editors, *Heuristics and biases: The psychology of intuitive judgment*, pages 1–18. Cambridge University Press.
Goldman, A. (1986). *Epistemology and cognition*. Harvard University Press.
Greco, D. (forthcoming). *Idelization in epistemology: A modest modeling approach*. Oxford University Press.
Hartmann, S. and Sprenger, J. (2010). Bayesian epistemology. In Pritchard, D. and Bernecker, S., editors, *The Routledge companion to epistemology*, pages 609–620. Routledge.
Hedden, B. (2012). Options and the subjective ought. *Philosophical Studies*, 158:343–360.
Holroyd, J. (2016). What do we want from a model of implicit cognition? *Proceedings of the Aristotelian Society*, 116(2):153–79.
Holyrod, J., Scaife, R., and Stafford, T. (2017). Responsibility for implicit bias. *Philosophy Compass*, 12:e12410.
Howes, A., Lewis, R., and Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, 116(4):717–51.
Howhy, J. (2013). *The predictive mind*. Oxford University Press.
Icard, T. (2018). Bayes, bounds, and rational analysis. *Philosophy of Science*, 85(1):79–101.
Johnson, E. and Payne, J. (1985). Effort and accuracy in choice. *Management Science*, 31(4):395–414.
Jones, M. and Love, B. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences*, 34:169–231.
Joyce, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.
Kahneman, D., Slovic, P., and Tversky, A., editors (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
Karlan, B. (2021). Reasoning with heuristics. *Ratio*, 34(2):100–8.
Langley, P., Laird, J., and Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10:141–60.
Lieder, F. and Griffiths, T. (2017). Strategy selection as rational metareasoning. *Psychological Review*, 124(6):762–94.
Lieder, F. and Griffiths, T. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:E1.
Lieder, F., Griffiths, T., Huys, Q., and Goodman, N. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin and Review*, 25(1):322–49.
Mandelbaum, E. (2016). Attitude, inference, association: On the propositioanl structure of implicit bias. *Nous*, 50(3):629–58.
Marewski, J., Gaissmaier, W., and Gigerenzer, G. (2010). We favor formal models of heuristics rather than lists of loose dichotomies: a reply to evans and over. *Cognitive Processing*, 11(2):177–9.
Morton, J. (2017). Reasoning under scarcity. *Australasian Journal of Philosophy*, 95(3):543–59.
Na'aman, O. (2021a). Emotions and process rationality. *Australasian Journal of Philosophy*, 99(3):531–46.
Na'aman, O. (2021b). The rationality of emotional change: Toward a process view. *Noûs*, 55(2):245–69.
Oaksford, M. and Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
Oswald, F., Gregory, M., Hart, B., James, J., and Tetlock, P. (2013). Predicting ethnical and racial discrimination: A meta-analysis of iat criterion studies. *Journal of Personality and Social Psychology*, 105(2):171–92.
Pasnau, R. (2013). Epistemology idealized. *Mind*, 122(488):987–1021.
Podgorski, A. (2017). Rational delay. *Philosophers' Imprint*, 17(5):1–19.
Polonioli, A. (2013). Blame it on the norm: The challenge from 'adaptive rationality'. *Philosophy of the Social Sciences*, 44(2):131–50.
Proust, J. (2013). *The philosophy of metacognition*. Oxford University Press.
Rescorla, M. (2019). A realist perspective on bayesian cognitive science. In Nes, A. and Chan, T., editors, *Inference and consciousness*, pages 40–73. Routledge.
Saul, J. (2013). Scepticism and implicit bias. *Disputatio*, 5(37):243–63.
Schooler, L. J. and Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112(3):610–628.
Simmons, A. J. (2010). Ideal and nonideal theory. *Philosophy and Public Affairs*, 38(1):5–36.
Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118.

Simon, H. (1976). From substantive to procedural rationality. In Kastelein, T., Kuipers, S., Nijenhuls, W., and Wagenaar, R., editors, *25 years of economic theory*, pages 65–86. Springer.
Simon, H. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41(1):1–20.
Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–90.
Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese*, 192(9):2769–93.
Staffel, J. (2020). *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.
Staffel, J. and De Bona, G. (2018). Why be (approximately) coherent? *Analysis*, 78(3):405–15.
Stewart, N., Chater, N., and Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, 53(1):1–26.
Stich, S. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. MIT Press.
Sylvan, K. (2020). An epistemic non-consequentialism. *Philosophical Review*, 129(1):1–51.
Thorstad, D. (2021). Inquiry and the epistemic. *Philosophical Studies*, 178(9):2913–28.
Thorstad, D. (forthcoming a). The accuracy-coherence tradeoff in cognition. *British Journal for the Philosophy of Science*, page forthcoming.
Thorstad, D. (forthcoming b). Two paradoxes of bounded rationality. *Philosophers' Imprint*, page forthcoming.
Thorstad, D. (forthcoming c). *Inquiry under bounds*. Oxford University Press.
Todd, P. and Gigerenzer, G. (2012). *Ecological rationality: Intelligence in the world*. Oxford University Press.
Viale, R. (2021). Why bounded rationality? In Viale, R., editor, *Routledge handbook of bounded rationality*, pages 1–54. Routledge.
Vul, E., Goodman, N., Griffiths, T. L., and Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, 38(4):599–637.
Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 14(1):101–18.
Wedgwood, R. (2018). Epistemic teleology: Synchronic and diachronic. In Ahlström-Vij, K. and Dunn, J., editors, *Epistemic consequentialism*, pages 85–112. Oxford University Press.
Wheeler, G. (2020). Less is more for bayesians, too. In Viale, R., editor, *Routledge handbook on bounded rationality*, pages 471–83. Routledge.
Wiens, D. (2020). The general theory of the second best is more general than you think. *Philosophers' Imprint*, 5:1–26.
Zhu, J.-Q., Sandborn, A., and Chater, N. (2020). The bayesian sampler: Generic bayesian inference causes incoherence in human probability judgments. *Psychological Review*, 127(5):719–48.

How to cite this article: Thorstad, D. (2024). Why bounded rationality (in epistemology)? *Philosophy and Phenomenological Research*, 108, 396–413. https://doi.org/10.1111/phpr.12978