Inquiry under bounds

David Thorstad
Comments welcome, but please ask before citing

Abstract

My aim in this paper is to develop an account of rational inquiry for bounded agents. I argue that an account should meet three minimal criteria. First, it should be tradeoff-sensitive, saying how scarce cognitive and non-cognitive resources are to be allocated within the course of a single inquiry, between competing inquiries, and between inquiry and other activities. Second, it should be stakes-sensitive, saying why more resources should typically be invested in our most important inquiries. And third, it should explain the irrationality of many instances of stereotyping, despite the superficial resemblance of stereotyping to cases of rational heuristic inquiry. I argue that we cannot meet these criteria by extending traditional accounts of rational belief or by applying existing philosophical accounts of bounded rationality. I develop a reason-responsiveness consequentialist view of rational inquiry for bounded agents and argue that this view fares well by the lights of the minimal criteria.

1 Introduction

Humans are bounded agents. We have limited cognitive abilities and must ration scarce cognitive and noncognitive resources during inquiry to achieve the best outcomes possible given our abilities. The study of bounded rationality asks what rationality demands of bounded agents.[1]

Herbert Simon held that a fundamental turn in the study of bounded rationality is the shift from substantive rationality to procedural rationality (Simon 1976). Substantive rationality asks normative questions about the outcomes of inquiry such as belief, intention, or preference. By contrast, procedural rationality asks normative questions about

---

[1]For recent philosophical work on bounded rationality see Icard (2018), Morton (2017), Wheeler (forthcoming) and relatedly Gigerenzer (forthcoming) and Schmidt (2019).

the processes of theoretical and practical inquiry by which these outcomes are produced.[2] Most theorists in the bounded tradition have followed Simon in thinking that normative theory should pay central attention to inquiry.[3]

My project in this paper is to develop an account of rational inquiry for bounded agents. My focus will be on theoretical inquiry, although much of what I say has implications for practical inquiry as well. In Section 2, I specify three minimal criteria for a successful account. In Sections 3-4, I draw lessons from two types of existing accounts while arguing that they fail to meet the minimal criteria in a satisfactory way. In Sections 5-6, I develop a new *reason-responsiveness consequentialist view* (RRCV), then in Sections 7-8 I show how the RRCV meets the minimal criteria.

2   Minimal criteria

In this section, I lay out three minimal criteria on an account of rational inquiry by bounded agents. Each criterion holds that normative theory should be sensitive to an important type of challenge faced by bounded inquirers. These criteria will guide our search for an account.

2.1   Tradeoff-sensitivity

The first minimal criterion is *tradeoff-sensitivity*. Inquiry, like most activities, takes place under conditions of scarcity. Candidate inquiries compete among themselves and with other activities for scarce cognitive resources such as memory, computational bandwidth, and executive control. Inquiries also often compete for non-cognitive resources such as time and money, and generate non-cognitive costs such as carbon emissions during travel.

---

[2]Simon, like many in the bounded tradition, adopted an indirect form of normative assessment on which the normative status of outcomes is dependent on the normative status of the inquiries that produced them. That is a separable claim which I will not take on board.

[3]This question is intimately related to the burgeoning philosophical literature on rational inquiry, insofar as central normative phenomena such as clutter avoidance (Friedman 2019; Harman 1986), question settling (Bratman 1987; Friedman forthcoming a) and logical non-omniscience (Harman 1986; Hoek forthcoming) are bounded rationality phenomena. I discuss implications for these questions in [removed].

These conditions of scarcity generate three types of tradeoffs that an account of rational inquiry should be sensitive to.

First, there are *intra-inquiry tradeoffs* within the course of a single inquiry. Most famously, there is often an *accuracy-effort tradeoff* in strategy selection (Johnson and Payne 1985).[4] More cognitively demanding strategies tend to produce more accurate judgments, so agents must select strategies which strike an appropriate balance between accuracy and effort. For example, we must choose how much information stored in memory to retrieve before making a judgment (Vul et al. 2014). Each additional item of information retrieved increases the expected accuracy of our judgment, but incurs a cognitive cost. Rational inquirers must balance the cognitive costs and benefits of information retrieval during inquiry. I return to this example in Section 7.

Second, there are *inter-inquiry tradeoffs* between different inquiries. We cannot inquire into all matters at once, so opening inquiry into one question incurs an opportunity cost of foregone inquiry into other questions. For example, while a Mars rover collects scientific information at one site it foregoes the opportunity to collect samples at another site (Zilberstein et al. 2001). Inter-inquiry tradeoffs have been a central feature of contemporary philosophical work on inquiry. For example, it is sometimes held that rational inquirers should construct a research agenda of questions to guide their inquiries (Olsson and Westlund 2006). Each item added to the research agenda leaves fewer resources to devote to other inquiries, so each additional question should be important enough to justify the opportunity costs that it imposes.

Finally, there are *inter-activity tradeoffs* between inquiry and other activities. Inquiry is one of many activities in a flourishing human life. Inquiry competes with other activities for resources such as time, money and attention. For example, every day we confront the following choice: whether to spend an extra hour in the library studying philosophy or to go home and have dinner with our families. A good normative theory should tell us how to make such choices by balancing the costs and benefits of inquiry against the costs

---

[4]Sometimes, there is no accuracy-effort tradeoff. In some situations, the most accurate strategies are also among the most frugal (Gigerenzer and Brighton 2009).

and benefits of competing activities.

Summing up, an account of rational inquiry for bounded agents should be tradeoff-sensitive, saying how tradeoffs should be made within a single inquiry, amongst several inquiries, and between inquiry and other activities.

2.2   Stakes-sensitivity

A second criterion of adequacy is *stakes-sensitivity*. A good theory should say how the stakes of inquiry bear on its rationality. For example, other things equal we should reserve the most frugal heuristic strategies for unimportant matters and use effortful non-heuristic strategies to make important judgments. And even once a strategy is selected, cognitive resources should be rationed according to the importance of the problem at hand. For example, we should retrieve more information from memory (Vul et al. 2014) and be more willing to gather new evidence from our environment (Stigler 1961) to solve the most important cognitive challenges.

To see the need for a stakes-sensitive theory, consider the psychology of poverty (Morton 2017). Poverty significantly impairs agents' performance on tasks measuring reasoning (Deck and Jahedi 2015; Mani et al. 2013), attention (Shah et al. 2012), memory (Evans and Schamberg 2009) and executive control (Mani et al. 2013; Vohs 2013). These impairments can be quite severe, equivalent to a full night spent without sleep (Linde and Bergström 1992; Mani et al. 2013). These cognitive impairments perpetuate poverty traps by contributing to behaviors such as overborrowing (Shah et al. 2012), undersaving (Bernheim et al. 2015), and noncompliance with medical instructions (Kaplan et al. 2004). Do these findings show that poverty breeds irrational inquiry? That seems an uncharitably mean-spirited conclusion to draw, but how can it be avoided?

To see what is going on here, suppose I ask you to imagine that your car has broken down and requires a $1,500 repair. Then I assign you a reasoning task. If you are financially well-off, your performance on the reasoning task will be unimpaired (Mani et al. 2013). But if you struggle financially, the story of the broken-down car dominates your cognition.

4

How will you pay for the repair? If you cannot pay, will you lose your job? Your home? Your ability to reason, attend, remember and control cognition in matters unrelated to the car will be substantially impaired because you are busy thinking about the broken-down car. But that is not irrational. You are correctly focusing on what matters most: the broken-down car.

Now we can see exactly how poverty impairs cognition.[5] Poverty creates a number of immediately pressing cognitive challenges, and agents respond by rationally reallocating the bulk of their cognitive resources towards these high-stakes challenges while neglecting the rest. Here is how a leading study puts the point.

> The human cognitive system has limited capacity. Preoccupations with pressing budgetary concerns leave fewer cognitive resources available to guide choice and action. Just as an air traffic controller focusing on a potential collision course is prone to neglect other planes in the air, the poor, when attending to monetary concerns, lose their capacity to give other problems full consideration. (Mani et al. 2013, p. 976).

The rational response to poverty is not a stakes-neutral allocation of cognitive resources, split equally between monetary concerns such as overdue electric bills and more trivial concerns such as making weekend plans. The rational response to poverty is a stakes-sensitive allocation of cognitive resources according to the importance of the cognitive challenges that agents face.

The stakes of cognition do not only matter to the poorest of us. All agents face cognitive challenges of varying importance. And all of us, if we are rational, allocate cognitive resources in a stakes-sensitive way, devoting the bulk of our resources to the most important cognitive problems. A good account of rational inquiry should say why this is so.

---

[5]There is another way in which poverty impairs cognition, namely by decreasing the stock of available cognitive resources through challenges such as sleep deprivation (Patel et al. 2010), stress (Cohen et al. 2006), and malnutrition (Gailliot et al. 2007) and increasing the total number of cognitive problems to be solved (Mani et al. 2013).

2.3 Explaining the irrationality of stereotyping

To meet the demands of a busy cognitive life, rational agents often make use of heuristics, cognitive strategies with low informational demands or with frugal rules for processing informational inputs (Gigerenzer and Gaissmaier 2011). For example, you might judge that one city is larger than another when you recall that the first, but not the second, has an intercity train line (Gigerenzer and Goldstein 1996), or judge that one tennis player will win a match because you have never heard of her opponent (Serwe and Frings 2006). Retrieving and processing information from memory is a costly activity, and in unimportant tasks or when time is of the essence, we often do well to make judgments based on small samples of information (Icard 2018; Vul et al. 2014). The resulting judgments are reliable enough to meet task demands without imposing undue cognitive costs.

Here is an uncomfortable fact. Stereotypes bear a striking similarity to rational heuristics. On some views, stereotyping just is a cognitive heuristic (Bodenhausen 1990). Suppose that a busy executive meets a woman in an office building and judges that she is a secretary, because most women in the building are secretaries. The executive uses a subset of available information, namely the woman's gender presentation, and processes that information according to a simple inference rule by which female-presenting individuals in the office building are judged to be secretaries. Consistent with the heuristic interpretation of stereotyping, agents rely increasingly on stereotypes under conditions of low motivation (Neuberg and Fiske 1987), high task complexity (Bodenhausen and Lichtenstein 1987) and high cognitive load (Gilbert and Hixton 1991), exactly the conditions that rationally drive us towards heuristic processing. We need to explain why many instances of stereotyping, such as the executive's inference, are impermissible, while similarly frugal and reliable inferences about cities and tennis matches are rationally permissible.

Here it is important but insufficient to criticize the beliefs or acceptances that result from stereotyping, for example on the grounds that they exhibit some evidential defect (Gardinier 2018). That is because rational belief is one matter, and rational inquiry an-

other.[6] Even if it is held that rational beliefs must be evidentially supported, we should not hold that rational inquiry requires examining all or even most relevant evidence held in memory, perception, or the agent's environment. As we have seen, rational inquirers often neglect most of the available evidence in order to balance the importance of accuracy and effort in cognition. A good account of rational inquiry needs to explain why rational inquirers can permissibly run a sizable risk of forming evidentially unsupported beliefs about cities and tennis players, but not about secretaries.

We have our marching orders. A good account of rational inquiry should be tradeoff- and stakes-sensitive, and explain the irrationality of many cases of inquiry by stereotyping. The task remaining is to develop an account of rational inquiry which meets these criteria. In Sections 3-4, I consider and reject two families of accounts. Then in Sections 5-6, I develop a new account, and in Sections 7-8 I show how this account meets the minimal criteria.

3   Extending theories of rational belief

A natural first try is to extend traditional theories of rational belief into theories of boundedly rational inquiry.[7] In this section, I argue that the first try fails for two reasons.[8] First, because theories of rational belief were not always developed with inquiry in mind, it is often unclear what these views could be extended to say about inquiry. Second, because theories of rational belief were not designed to meet the minimal criteria, natural extensions of traditional theories will fail to satisfy the minimal criteria. This is no fault of existing theories of rational belief, because they were not intended as accounts of boundedly rational inquiry.

To see both problems in context, begin with a simple version of process reliabilism. On

---

[6]I expand on this point in the discussion of level separation in Section 6.

[7]Why not use existing theories of rational inquiry from the recent literature on zetetic epistemology? As yet, this literature has not produced a complete theory of rational inquiry. Elsewhere [removed], I argue that the RRCV is a promising candidate.

[8]Nearly the same argument is made by Jane Friedman (2019; forthcoming a; forthcoming b).

this view, a belief is rational just in case it is the output of a sufficiently reliable process. The first problem is that it is unclear how to extend reliabilism to account for the rationality of processes which do not immediately output beliefs. Does reliabilism tell agents how to allocate cognitive resources among competing inquiries, for example inquiries into an overdue electric bill and a broken-down car? And what does reliabilism have to say about the questions of when and how agents should gather evidence? There are certainly proposals which come to mind, but none which bears an especially tight relationship to process reliabilism in its contents, motivations, or grounds.

The second problem is that when it is reasonably clear how to extend reliabilism, the extended theory does not meet the minimal criteria. For example, consider the selection of inference rules (Lieder and Griffiths 2017; Marewski and Schooler 2011). Suppose I am trying to figure out whether Jones is angry with me. I'd like to decide whether to infer heuristically or nonheursitically, and more specifically which inference rule to use. A natural reliabilist theory says that I may use any sufficiently reliable inference rule.

But this theory satisfies none of the minimal criteria. First, reliabilism is not stakes-sensitive. It says that I may reason in the same way if Jones is my partner or a total stranger. Second, reliabilism has little to say about inter-inquiry or inter-activity tradeoffs.[9] Reliabilism does not tell me how to balance my inquiry into Jones' mental state against my philosophical research or the need for a good night's sleep. Third, reliabilism does not explain the irrationality of stereotyping. A key datum in recent discussions of rational stereotyping is that stereotyping can be irrational even when equally reliable inference rules would be rational in other contexts (Basu 2019). But our extended reliabilist theory denies this.

Of course, we could help reliabilism to meet the minimal criteria by saying that rational inference rules should meet a context-sensitive reliability threshold and letting stakes, tradeoffs, and the importance of avoiding bias impact this threshold. But now it is the theory of threshold-setting and not reliabilism itself carrying the weight of the account.

---

[9]In fact, tradeoff-insensitivity has recently been taken to be an advantage of the reliabilist view (Ahlstrom-Vij and Dunn 2014).

And that theory will look a lot like the theories considered in the rest of this paper.

Similar remarks apply to other epistemological views such as evidentialism and coherentism.[10] It is not always clear how to extend these views to answer questions about rational inquiry. And when we try, the resulting extensions fail to meet the minimal criteria. The lesson is that boundedly rational inquiry is a new subject matter which deserves a new account.

4   Existing theories of bounded rationality

I am not the first philosopher to develop an account of bounded rationality. In this section, I review three existing accounts. The first two, I argue, may well be true, but neither provides a complete theory of boundedly rational inquiry. The third is false, but points the way towards the correct account.

One prominent account of bounded rationality is Christopher Cherniak's minimal rationality (Cherniak 1981, 1986). The minimal rationality program weakens standard rationality requires to accommodate agents' bounds. So for example, instead of the

> **(Ideal General Rationality Condition)** If an agent has a particular belief-desire set, he would undertake all actions which are apparently appropriate. (Cherniak 1981, p. 164).

minimally rational agents satisfy a weaker condition:

> **(Minimal General Rationality Condition)** If an agent has a particular belief-desire set, he would attempt some, but not necessarily all, of those actions which are apparently appropriate. (Cherniak 1981, p. 166).

This account may well be correct, but it is twice removed from our project.

---

[10]A cousin of coherentism has recently been defended as a theory of bounded rationality by Staffel (2020). This view deserves separate treatment; see [removed] and (Arkes et al. 2016).

The first removal is that minimal rationality conditions are requirements of substantive rationality. They say what rational agents will do, believe, intend and the like. We saw in the last section that requirements of substantive rationality cannot easily be extended into theories of procedural rationality, which say how agents should inquire. Philosophers studying minimal rationality have had illuminating things to say about processes of inquiry such as memory search, but these remarks are largely separable from the minimal rationality conditions, which are themselves not conditions on processes of inquiry.

The second removal draws on a distinction between two theoretical roles that an account of rationality can play (Bermúdez 2009). The first is to predict and explain agents' behavior. The second is to normatively assess behavior. Theories of minimal rationality have been primarily developed as predictive-explanatory theories.[11] They aim to specify the minimal conditions under which it makes sense to describe agents using intentional vocabulary. And these conditions can be quite permissive. Cognitive scientists are increasingly willing to attribute intentional states to lower animals and even perhaps to plants (Adams 2018).

The predictive-explanatory question of whether it makes sense to attribute intentional states to an agent is separable from the normative question of how her inquiries should be assessed. Being the bearer of intentional attitudes is a precondition for normative evaluability. But agents may be rational in the predictive-explanatory sense of bearing intentional attitudes while being irrational in the sense that some of those token attitudes warrant negative normative assessments. In this way, standard accounts of minimal rationality are twice removed from theories of boundedly rational inquiry. They ask when it makes sense to attribute intentional states to agents, rather than how we should normatively assess the inquiries of agents who have intentional states.

A second account of bounded rationality is due to Edward Stein (1996). Stein defends what he calls a naturalistic account of rationality on which rationality is determined by

---

[11] For example, Cherniak (1981, p. 161) begins with the following question: "In intensional explanations of behavior, we require rationality of the agent. How rational must a creature be to be an agent, that is, to qualify as having a cognitive system of beliefs, desires, and perceptions?" An exception to this predictive-explanatory interpretation of minimal rationality is Hoek (forthcoming).

reflective equilibrium on four types of inputs: first-order judgments about the nature of good reasoning; intuitions about abstract principles of reasoning; philosophical theories; and scientific theories and data. Stein uses this naturalistic account to show how empirical results can motivate scientifically driven revisions to philosophical theories about rationality, often vindicating the rationality of purported cognitive biases.

The point is well taken, and I will consider a specific instance of this vindicatory project in Section 7. But Stein's naturalistic account is, perhaps by design, not specific enough to pin down any particular theory of boundedly rational inquiry. Most normative theories including my own will claim to occupy the best reflective equilibrium between Stein's four inputs. As a result, we can take on board Stein's naturalistic account without halting our search for a more specific theory.

A third account of bounded rationality is due to Steven Stich. Stich sketches a pragmatic account, offering the following tentative statement of his account:

> In evaluating systems of cognitive processes, the system to be preferred is the one that would be most likely to achieve those things that are intrinsically valued by the person whose interests are relevant to the purposes of the evaluation. In most cases, the relevant person will be the one who is or might be using the system. (Stich 1990, p. 131).

Pragmatic views are also popular in the scientific literature on bounded rationality.[12]

Stich's pragmatism is twice removed from my own project. First, it is offered as an account of how cognitive systems should be evaluated, whereas my goal is to evaluate token inquiry processes. Second, Stich's pragmatism is phrased at a high level which leaves open whether it is a claim about rationality, rightness, virtue, or another normative category. But what would happen if we repurposed Stich's view as an account of the rationality of token inquiry processes?

---

[12]For example, Peter Todd and coauthors maintain that: "The success of simple heuristics is defined with respect to pragmatic goals in a particular environmental context." (Todd et al. 2000, p. 378).

On this pragmatist view, a process of inquiry is permissible just in case it is as conducive as any alternative to what the relevant agent intrinsically values. This repurposed pragmatist view does meet the minimal criteria. Pragmatism is tradeoff-sensitive, instructing agents to make tradeoffs within and between inquiries and other activities according to the relevant agent's intrinsic values. Pragmatism is stakes-sensitive, with the stakes read directly off from intrinsic values. And pragmatism says that stereotyping is irrational when the relevant agent places sufficiently high disvalue on bias and discrimination.

However, there are two worries about the way in which pragmatism meets the minimal criteria. First, a soft worry: pragmatism makes questions about rational cognition highly indeterminate. The intrinsic values of bounded agents are underdetermined, often radically so. This happens because it is not cognitively efficient to attach precise weights to valuable quantities such as family, health, career success, and environmental preservation. On many leading decision theories, indeterminacy in agents' values generates substantial indeterminacy in what it is rational for them to do (Troffaes 2007). Bounded agents will be substantially undecided about how to weigh our fundamental values, and this indeterminacy propagates downstream throughout many questions about rational inquiry. For a pragmatist, there may be no answer to whether I should stay in the library for an extra hour or go home to my family until I make up my mind about what I truly value.

Many philosophers will be comfortable admitting some indeterminacy in rational evaluation, but pragmatism faces a tougher worry. As Stich recognizes, pragmatism relativizes the normative status of inquiries to the values that agents hold. Consider first what happens if the person whose interests are relevant is always the agent acting, as in traditional pragmatist theories. This theory says the wrong thing about the rationality of stereotyping. It says that stereotyping is irrational only if the actor cares about avoiding bias. A midcentury executive who is mostly unconcerned with the welfare of women would be rationally required to stereotype them in order to advance his interests and avoid excessive cognitive labor. Likewise, this pragmatist theory answers questions about

stakes and tradeoffs by exempting an actor's intrinsic values from rational criticism. If I genuinely value a life of debauchery, then I cannot be criticized for the damage that I do to my own well-being or the well-being of others except on the grounds that I have been insufficiently debaucherous.

Pragmatists have two routes of reply here. One strategy is to hold that pragmatism captures a minimal Humean or instrumentalist notion of rationality on which intrinsic values are not open to rational evaluation. If the pragmatist is willing to accept that there is another sense of rationality which is non-Humean, or that questions about what we ought to do are non-Humean, then she is invited to read my account in that way.[13] Otherwise, I am happy to conditionalize my view on the success of anti-Humean arguments (Korsgaard 1997; Lavin 2004).

A second strategy is to shift the relevant agent whose interests matter rationally. Stich allows that the agent whose interests matter to normative evaluations may not always be the actor. So for example, Stich could hold that stereotyping is irrational because it threatens the interests of stereotyped individuals. But such a theory, if the details could be filled out, would take us well beyond traditional versions of pragmatism and towards competing theories. This is particularly true if Stich allows, as I think he should, that the interests of multiple agents may matter to a given normative evaluation. Alternatively, it might be that Stich's project here is to capture a type of assessor-relativism on which the relevant interests are the interests of the agent making a normative judgment.[14] If so, this is another point at which I am comfortable parting ways.

In this section we met three philosophical accounts of bounded rationality. The first two may well be true, but do not provide complete theories of bounded rationality. The third is promising but has two undesirable features: relativism and indeterminacy. Can we construct an account of boundedly rational inquiry which removes these features, while preserving the key insights of a pragmatist account? I take up this project in the

---

[13]This strategy may be in tension with some of the claims made in Section 5, but need not be incompatible with the core of the account in Section 6.

[14]For suggestive remarks see (Stich 1990, pp. 134-45).

next two sections.

5 Reason-responsiveness

A good theory of rationality should tell us how rationality relates to reasons. To get these issues into focus, it will help to modify one of our initial examples.

Imagine that you are living paycheck-to-paycheck. You come home late from work intending to sort through the monthly bills and figure out how to pay off as many bills as you can. You hear a noise outside and rush out to discover that someone has backed into your car and driven off. Now the car will not start. You don't know how you will fix the car or what you will do for transportation in the meantime.

You do not have enough time or mental energy to think effectively about how to pay your bills and also how to fix your car. You need to choose which problem to focus on tonight. If you are rational, will you allocate your attention to the car or to the bills? Plausibly, you are rationally required to think about the car and not the bills. A natural explanation for this requirement is that you have more reason to think about the car than to think about the electric bills. Rationality requires you to do what you have most reason to do.

But doing what you have most reason to do is not sufficient for inquiring rationally. Suppose that you focus your reasoning on the car. But you do not do this because you are concerned about getting to work or fixing the car. You do this because you are convinced that thinking about the car will please the ghost of your father, who always loved cars. Plausibly, your inquiry is not rational. A natural explanation for this fact is that inquiring rationally requires not only doing what you have most reason to do, but also acting in response to the reasons which make your action rational. If you had focused on the car in response to the need to secure transportation, that inquiry would have been rational. But it is not rational to focus on the car to appease a paternal ghost.

These thoughts are often taken to motivate a reason-responsiveness conception of

rationality on which rationality requires responding correctly to reasons. More precisely:

> **(Rationality as Reason-Responsiveness)** For all agents S, times t and acts X,
> S's X-ing at t is rational if and only if in X-ing at t, S responds correctly to the
> normative reasons that she possesses at t.

When thinking about bounded agents, it is important to be generous in our interpretation of the responsiveness relation. Responding to the need for transportation does not require explicitly considering the question of how you will get to work tomorrow and making the conscious inference that you should think about how to fix your car. Leading accounts of the responsiveness relation hold that agents respond to reasons when those reasons play an appropriate causal role in producing action (Arpaly and Schroeder 2015) or produce action through a mediating competence to respond to reasons (Lord 2018). These notions of reason-responsiveness allow bounded agents to respond to reasons without imposing undue cognitive burdens that bounded agents cannot efficiently meet.

The reason-responsiveness conception of rationality has two important consequences. First, *rationality is deontic:* what we are rationally required to do and what we ought to do are the same thing (Lord 2017). That is:

> **(Rationality is Deontic)** For all agents S, times t and acts X, S is rationally
> required at t to X just in case S ought, at t, to X.

Roughly put, the claim that rationality is deontic follows from a reason-responsiveness view because what S ought to do and what she is rationally required to do are determined in the same way: by the balance of possessed normative reasons.[15]

The claim that rationality is deontic allows us to explain why epistemologists often pass between claims about what agents ought to do and what they are rationally required to do. On a reason-responsiveness view this passage is not a conceptual mistake, because rational requirements and oughts are coextensive. At the same time, the reason-responsiveness view explains why rationality and rightness remain importantly separate

---

[15]For a full argument see Kiesewetter (2017) and Lord (2018).

normative categories. From an ex ante perspective, oughts and rationality coincide: we are rationally required to token all and only those acts we ought to token. But from an ex post perspective, rationality is a strictly more demanding status. For *S*'s token act of *X*-ing to be rational requires not only that *S* ought to have *X*'ed, but also that *S X*'ed in response to the reasons she possessed for *X*-ing.

A second consequence of the reason-responsiveness conception is that *rationality is all-things-considered*.[16] By this, I mean that all types of normative reasons and not just a privileged subclass of prudential or epistemic reasons bear on the rationality of agents' actions. This consequence is important, because it allows us to see why theories such as reliabilism and pragmatism are not suitable accounts of rational inquiry for bounded agents. Reliabilism privileges a class of truth-directed intellectual reasons, whereas pragmatism privileges pragmatic reasons for agents to promote their own goals. While there may be some contexts in which these are the only reasons at issue, inquiry is not such a context. The minimal criteria stress that rational inquirers are concerned not only for truth or prudential gain, but also to balance a wide array of stakes and tradeoffs in order to promote values such as welfare and nondiscrimination for themselves and others. The reason-responsiveness position is better-placed to incorporate the diverse range of reasons bearing on rational inquiry because it takes an all-things-considered stance from which all types of normative reasons are relevant.

These consequences of the reason-responsiveness conception can also be independently motivated. Consider *Kolodny's challenge*: why be rational (Kolodny 2005)? The reason-responsiveness view answers Kolodny's challenge in the natural way: in acting rationally we do what we ought to do and also what we have most reason to do. But on many views which deny that rationality is deontic or all-things-considered, there can be acts *X* and *X'* such that I have more reason to *X* or ought to *X*, but am nonetheless rationally required to *X'*. Then Kolodny's challenge sharpens to the following: why be rational,

---

[16]It is an important question whether there is also a separate notion of epistemically rational inquiry, and whether this notion deserves special privilege in theorizing. In [removed] I answer 'yes' to the first question, and 'no' to the second.

instead of doing what I ought or have most reason to do? These questions are difficult to answer in a way that preserves the authority of rationality, and the reason-responsiveness conception allows us to avoid them.

In this section, we have developed and motivated a reason-responsiveness conception of rationality along with two of its consequences: the claims that rationality is deontic and all-things-considered. This takes us halfway towards an account of rational inquiry. To go further, we will need to say something specific about the reasons that bounded agents face.

6   Reasons for inquiry

In the last section, we developed a reason-responsiveness view of rational inquiry. It follows from this view that agents are rationally required to inquire in the ways that they ought and have most reason to inquire. To go further, we will have to adopt some specific views about normative reasons for inquiry.

Many views are possible here. In this section, I develop a reason-responsiveness consequentialist view (RRCV) which extends the reason-responsiveness account in consequentialist fashion. In developing the RRCV I will assume a background consequentialist view and ask how that view could be developed to account for boundedly rational inquiry. This background view will need to be motivated in traditional ways (Parfit 1984; Portmore 2011). In this section, I show how the RRCV could be developed and in the next two sections, I show how the RRCV meets the minimal criteria. It would be an interesting project to repeat the analysis in these last three sections using an alternative view.

Like any consequentialist view, the RRCV will have to hold that agents have most reason to do what is best.[17] That is:

> **(Reason-Evaluative Bridge)** For all agents S, times t and acts X, S has most

---

[17]That is, unless we adopt a satisficing consequentialist view (Slote 1984). However, this view is not forced on us by considerations of bounded rationality (Pettit 1984) and faces challenges (Bradley 2006; Byron 1998; Mulgan 2001).

reason at t to X just in case no alternative to X at t for S is better than X.[18]

From this principle, we can derive a traditional consequentialist deontic theory using the claim that agents ought to do what they have most reason to do.

> **(Ought-Reason Bridge)** For all agents S, times t and acts X, S ought, at t, to X just in case S has most reason at t to X.

Combining these principles, we recover the first part of a traditional consequentialist deontic theory:

> **(Ought-Evaluative Bridge)** For all agents S, times t and acts X, S ought, at t, to X just in case S's X-ing at t is best.

Here the phrase 'S's X-ing at t is best' is introduced as shorthand for the claim that there is no alternative to X at t for S which is better than X.

Because Rationality is Deontic, we can restate the Ought-Evaluative Bridge as a claim about rational requirements:

> **(Rationality-Evaluative Bridge)** For all agents S, times t and acts X, S is rationally required at t to X just in case S's X-ing at t is best.

The Rationality-Evaluative Bridge can be viewed as part of a global consequentialist approach to rational inquiry, insofar as it generalizes the traditional consequentialist view of right action to cover rational requirements on inquiry.[19]

But the Rationality-Evaluative Bridge is not yet a full consequentialist theory. Non-consequentialists can accept that we ought to inquire in the ways that are best (Sylvan 2020). They need only hold that what is best is sometimes to honor or promote, rather

---

[18]Readers who admit phenomena such as parity or incomparability of value may want to weaken the Reason-Evaluative Bridge. These readers are invited to make their preferred weakening and push through to generate a corresponding weakening of the Rationality-Evaluative Bridge.

[19]On global consequentialism see Driver (2012), Kagan (2000), Parfit (1984) and Pettit and Smith (2000).

than instantiate value.[20] For example, a Kantian might hold that it is irrational to make inferences based on racial stereotypes because in doing so we dishonor the truth as well as the targets of our stereotypes. By contrast, a consequentialist might hold that it is irrational to make inferences based on stereotypes because in doing so we promote false belief and harm others. Here the debate between consequentialists and their nonconsequentialist opponents concerns the question of whether all normative reasons are reasons to promote value. The RRCV supplements the Rationality-Evaluative Bridge with the claim that all reasons are reasons to promote value.

   (Promotion) All normative reasons are reasons to promote value.

Together with the Rationality-Evaluative Bridge, this implies that rational inquiry involves taking the inquiries which best promote value, or refraining from inquiry altogether if some competing activity better promotes value.

   One way to see the attraction of the RRCV to is consider how the view improves on pragmatism. Consequentialists think that the pragmatist has put the cart before the horse.[21] Pragmatism takes an agent's values as settled, beginning with the question of how agents should act given their values and answering that agents should act to promote their intrinsic values. By contrast, consequentialism begins with fundamental reflection on what matters, using this reflection to derive a theory of value and then holding that agents should act to promote what is actually of value. This means that unlike pragmatism, the RRCV does not relativize rational requirements to an agent's values. Nor does the RRCV say that rational requirements are indeterminate until agents have made up their minds about what they value. We can build both of these features into the RRCV if desired. For example, we can recover pragmatism by adopting a desire-satisfaction theory of value on which what matters is derived from the relevant agent's intrinsic desires. But the RRCV

---

[20]Nonconsequentialists may also deny the grounding claim that *X*-ing is rationally required because *X*-ing is best. Like many consequentialists, I accept that grounding claim but I will not adopt or argue for the claim here. This is important, because nonconsequentialists may still think that rational inquiries are typically among the most valuable even if they offer a different explanation for why this is so.

[21]Strictly speaking, the charge of putting the cart before the horse assumes that consequentialism is a claim about the grounds of normative facts.

also gives us the flexibility to avoid relativism and indeterminacy, as well as to subject agents' intrinsic values to criticism, by adopting a different axiology.

For the RRCV to be a plausible account of boundedly rational inquiry, we will need to introduce three more components. The first is a *rich axiology* on which intellectual values such as truth, knowledge and understanding carry significant final value alongside more traditional values such as equality and welfare. This will allow us to recover truth-directed norms of inquiry, as well as to explain the value of inquiries such as a life spent studying philosophy, much of whose value lies in the fact that it promotes intellectual goods. But by admitting non-intellectual values into our axiology, we will still be able to explain why even the most devoted philosopher should sometimes go home from the library and figure out how to fix their car.

Second, the RRCV adopts a sharp *level separation* between questions about rational belief and rational inquiry. In the practical domain, consequentialists traditionally distinguish between the questions of what an agent is rationally required to intend and how she is rationally required to deliberate about how to intend (Parfit 1984; Railton 1984). In the theoretical domain, consequentialists should draw an equally sharp distinction between questions about rational belief and rational inquiry. It is one thing to ask what agents are rationally required to believe and another thing entirely to ask how they should inquire. This will allow us to hold that both intellectual and non-intellectual values bear on rational inquiry without implying the revisionary claim that non-intellectual values bear on rational belief. While we are free to make that claim if desired, it should not be forced on us merely by reflection on the fact that rational inquiry answers to non-intellectual sources of value.

Third, in applying the account we will have to say something about how the values of actions such as inquiry are determined by the values of their possible outcomes.[22] *Objectivism* values actions at the values of their actual outcomes; *subjectivism* values actions at

---

[22]This presentation blurs two questions: (i) *which* body of information bears on the value of actions, and (ii) *how* that information bears on the value of actions. I assume that the answer to (ii) is given by expected utility theory. I suspect that many of the conclusions of this paper could be broadened to accommodate other decision rules.

the expected values of their actions given an agent's beliefs, and *information-sensitive* views value actions at their expected values given the relevant information or evidence.[23] I favor an information-sensitive view, although much of what I have to say will be compatible with subjectivism if the agents in question hold evidentially supported beliefs.[24]

However, many of the remarks made in the next two sections will be incompatible with objectivism. That is to be expected in a theory of bounded rationality. Many of our most important cognitive bounds are informational bounds. We do not know the outcomes that our actions will have. As a result, theories of bounded rationality need to evaluate actions in a way that incorporates agents' uncertainty about the outcomes that their actions will have. Objectivism does not do this.

In this section and the last, we have developed a reason-responsiveness, consequentialist view (RRCV) of rational inquiry. On the RRCV, agents are rationally required to inquire in the ways that they ought and have most reason to inquire. This in turn means that agents are rationally required to inquire in the ways that best promote value. The RRCV adopts a rich axiology on which both intellectual and non-intellectual outcomes bear final value; uses a sharp level-separation to restrict its verdicts to the case of rational inquiry without any direct consequences for the rationality of belief; and adopts an information-sensitive account on which inquiries are valued at their expected values given the relevant evidence.

It is time to test the RRCV. In the next two sections, I show how the RRCV meets the minimal criteria in a way that grounds plausible and illuminating explanations of the normative data.

---

[23]I assume here that the relevant body of information is the agent's total evidence, although this view could be broadened, for example to accommodate the evidence that the agent should have had. See Goldberg (2016).

[24]For arguments in favor of the information-sensitive view see Cariani (2016) and Kolodny and MacFarlane (2010) among many others.

7   Stakes and tradeoffs

In this section, I show how the RRCV meets the first two minimal criteria: tradeoff-sensitivity and stakes-sensitivity. Then in the next section I consider what the RRCV has to say about stereotyping.

The RRCV has a rich and explanatorily powerful story about how tradeoffs should be made: by considering the expected values of candidate inquiries. This gives a uniform account of intra-inquiry, inter-inquiry, and inter-activity tradeoffs. For example, we should choose between studying philosophy and counting blades of grass based on the expected value promoted by each activity, both the intellectual value of true belief and knowledge as well as the non-intellectual value of that knowledge in guiding action. And we should choose between studying philosophy and eating dinner with our families in exactly the same way. Consequentialism is also stakes-sensitive, since the stakes of inquiry are built directly into the expected value of inquiry. Ceteris paribus, we should invest more effort into important inquiries than unimportant inquiries because it is more important to be correct in important matters than in unimportant matters.

To see that the RRCV correctly captures the normative relevance of stakes and tradeoffs, we should look towards applications where the RRCV makes novel and correct normative predictions. In the rest of this section, I consider a case study designed to illustrate the strength and plausibility of the story that the RRCV tells about stakes and tradeoffs.

Suppose I ask you to estimate the year in which George Washington was first elected president. If you are like many people, you will begin by considering an *anchor value*: the year 1776 in which the Declaration of Independence was signed. You will then *adjust* your estimate upwards and downwards by considering new items of information. For example, you may know that the Revolutionary War lasted seven years, taking us to 1783. And if you are like most people, you will make several further adjustments, ending up with an estimate in the low- or mid-1780s (Epley and Gilovich 2006; Lieder et al. 2018).

This procedure is called *anchoring and adjustment*, and it is one of the three original

heuristics proposed by Tversky and Kahneman (1974). Anchoring and adjustment is often an effective way to make accurate judgments at low cognitive cost. For example, George Washington was first elected in 1788. But anchoring and adjustment exhibits a characteristic anchoring bias: judgments made by anchoring and adjustment tend to be biased towards the anchor. In this case, they are several years too low. Does this mean that anchoring and adjustment is an irrational method of inquiry? That would be too hasty, since most heuristics exhibit biases, but in many circumstances rational agents must rely on heuristics to balance the costs and quality of cognition. But how are we to tell whether and in what circumstances anchoring bias is a sign of irrational cognition?

The RRCV delivers a precise and plausible story about the circumstances under which anchoring bias emerges from agents' rational response to stakes and tradeoffs (Lieder et al. 2018). The task in anchoring and adjustment is to construct an estimate $\hat{x}$ of an empirical quantity, such as the year that George Washington was elected president. Agents begin with an initial estimate $\hat{x}_0$ provided by the anchor value. They then adjust this estimate by sampling information from memory and incorporating that information to generate revised estimates $\hat{x}_1, \hat{x}_2, \ldots$ [25]

Rational anchoring and adjustment responds to two separate costs. On the one hand, sampling information from memory incurs a cost of computation $cost(t)$, increasing with the number t of samples drawn.[26] On the other hand, sampling information decreases the error cost $cost(\hat{x})$ because it tends to produce a more accurate estimate.[27] Using a rich axiology, we can understand this error cost as a combination of the intrinsic value of forming true beliefs together with their instrumental importance in guiding action. Rational anchoring and adjustment involves selecting the optimal number t^* of adjustments to minimize the expectation of these combined costs.

---

[25]Lieder and colleagues model sampling using the Metropolis-Hastings algorithm (Hastings 1970; Metropolis et al. 1953) with a Poisson proposal distribution.

[26]Lieder and colleagues set $cost(t) = \gamma t$ for constant γ. For alternatives see (Shenhav et al. 2017).

[27]Lieder and colleagues set $cost(\hat{x}) = |\hat{x} - x|$ where x is the true value of the target variable. For alternatives see Gneiting and Raftery (2007).

$$t^* = argmin_t \mathbb{E}[cost(\hat{x}_t) + cost(t)].$$

By information-sensitivity, this expectation can be taken relative to the agent's total evidence. What light can this model shed on the rationality of anchoring bias?

An immediate prediction is that rational anchoring and adjustment will typically exhibit some degree of anchoring bias. The effect of the anchor on the final estimate washes out only as agents make a large number of additional adjustments. Each adjustment incurs a fixed cognitive cost but produces diminishing returns to the accuracy of agents' judgments, hence rational anchoring and adjustment typically halts before the anchoring bias can be eliminated. More generally, this model can be used to explain the rationality of patterns of anchoring bias exhibited by human agents. For example, anchoring bias increases under cognitive load as agents become busy with another task (Epley and Gilovich 2006). This is rationalized by the RRCV as an instance of inter-activity tradeoffs: increasing cognitive load increases the relative cost of computation by introducing a second inquiry which competes with the first for a shared stock of computational resources. The model can also explain why monetary incentives often reduce anchoring bias (Simmons et al. 2010) by invoking stakes-sensitivity. Monetary incentives raise the error cost of an incorrect judgment, increasing the rational number of adjustments that will be made and thereby decreasing the effect of the anchor.

These are simple and intuitive predictions about rational anchoring and adjustment which will be difficult for competing theories to make. Except for Stich's pragmatism, I know of no theory of rational inquiry in the contemporary philosophical literature which makes these predictions. And we have seen several grounds to prefer the RRCV to pragmatism.

More generally, there is a robust explanatory argument to be made for the RRCV on the basis of its ability to correctly describe rational patterns of cognition in tasks such as the selection of cognitive strategies (Lieder and Griffiths 2017), allocation of attention (Sims 2003); rational planning (Callaway et al. 2018); and the formation of good cognitive habits

(Kermati et al. 2016). In these applications as elsewhere, consequentialism tells a precise and plausible story about the normative relevance of stakes and tradeoffs, then uses that story to make novel predictions about the rationality of cognitive processes that might otherwise have appeared irrational. It is on the success of these explanatory accounts that the RRCV asks to be judged.

8   Consequentialism and stereotyping

Some inquiries are impermissible, despite their reliability, because they make unacceptable use of stereotypes during inquiry. Consider, for example:

> **(Secretary)** A busy executive during the 1960s passes a woman in the lobby of an office building. Knowing that over 95% of female-presenting employees are secretaries, the executive judges that the woman is a secretary.

Here we would like to say that the executive inquires impermissibly because this pattern of inquiry runs an unacceptably high risk of wrongly concluding that a female-presenting executive is a secretary. A puzzling feature of cases such as Secretary is that the executive's inquiry becomes permissible if the moral stakes are lowered.

> **(Employee)** A busy executive during the 1960s passes a person in the lobby of an office building. Knowing that over 95% of people in the building are company employees, the executive judges that the person is an employee.

Here the executive's inquiry may be rationally permissible. What explains this difference in permissibility?

A good start is Sarah Moss's argument that inquirers are subject to a Rule of Consideration:

> **(Rule of Consideration)** In many situations where you are forming beliefs about a person, you morally should keep in mind the possibility that they might be an exception to statistical generalizations. (Moss 2018, p. 221).

Moss holds that the Rule of Consideration applies in Secretary, but not Employee, hence in Secretary the executive wrongly fails to keep in mind the possibility that the employee might be an exception to the statistical generalization that most female-presenting employees are secretaries.

This explanation of the difference between Secretary and Employee raises two questions. The first concerns the scope and justification of the Rule of Consideration: when and why does this rule apply? The second concerns additional requirements beyond the Rule of Consideration: what else are rational inquirers required to do in order to avoid bias during inquiry? Let us tackle each question in turn.

The key insight regarding the scope and justification of the Rule of Consideration is provided by Renée Bolinger (forthcoming). The possibilities that we should keep in mind are determined not only by the strength of our evidence, but also by the seriousness of the harms that we risk by ruling out possibilities. We are required to keep in mind the possibility that a person in the lobby may fail to be a secretary but not the possibility that they may fail to be an employee, because the harms risked in Secretary far exceed those at stake in Employee. Gender discrimination during the 1960s led to significant deprivation of professional and social status as well as fair pay from women within the workforce, and contributed to pervasive forms of marginalization and exclusion throughout all other walks of life. By inferring that a female-presenting person in an office lobby is a secretary, the executive contributes to these harmful patterns of marginalization and exclusion.

What the RRCV adds to Bolinger's discussion is an explanation of how the magnitudes of risked harms and the probability of imposing them contribute to the irrationality of stereotyping.[28] Stereotyping becomes irrational when another, more careful pattern of inquiry has higher expected value, because the reduction in risked harms outweighs the cognitive and physical cost of more careful inquiry. A strength of this explanation is that the RRCV need not only be concerned with the harms that will arise if the executive's

---

[28]For discussions of how views similar to the RRCV should handle stereotyping, see Rinard (2019), Thomsen (2011) and the exchange between Risse and Zeckhauser (Risse and Zeckhauser 2004; Risse 2007) and Lever (Lever 2005, 2007).

conclusion is incorrect. One harm that the executive risks is the direct social and professional harm to female-presenting employees who are wrongly judged to be secretaries. But another harm that the executive risks is normalizing patterns of inference in which female-presenting employees are assumed, by default, to be secretaries. The RRCV can explain the relevance and importance of both types of harm to rational inquiry.

It might be objected that the RRCV is too soft on stereotyping, because it allows that the badness of stereotyping, like any other harm, can be outweighed. But this objection underestimates the magnitude of the harms inflicted by stereotyping. For but a few examples, prejudicial stereotypes fuel economic discrimination in employment (Bertrand and Mullainathan 2004), housing (Pager and Shepherd 2008) and credit markets (Munnell et al. 1996). In law enforcement, stereotypes produce dramatic disparities in rates of police search (Pierson et al. forthcoming), police violence (Zack 2015), imprisonment (Pettit and Western 2004) and harshness of sentencing (Steffensmeier and Demuth 2000). Experienced discrimination has devastating effects across measures of psychological (Schmitt et al. 2014) and physical (Pascoe and Richman 2009) well-being. Stereotypes produce epistemic costs such as lost confidence and knowledge in all parties involved (Gendler 2011). And beliefs based on stereotypes may themselves harm, for example by falsely diminishing their targets (Schroeder 2018).

Like many philosophers, I am not an absolutist. I think that any harm, however severe, can in principle be outweighed. But the harms imposed in many cases of stereotyping are quite severe and hence difficult to outweigh.

This explanation of the irrationality of stereotyping allows us to see why the superficial resemblance of stereotyping to cases of rational heuristic inquiry is misleading. Frugal heuristics are appropriate in low-stakes situations where the costs of cognition outweigh the importance of forming accurate beliefs. In Secretary, the executive wrongly confuses a passing interaction for a low-stakes situation because the executive fails to anticipate the magnitude of harm that his inquiry risks imposing. The executive's mistake is the same mistake that he would make if he judged that one player was likely to win a tennis

match because he did not recognize the other player, then signed the first player to a major sponsorship contract. Given the stakes of the executive's inquiry, he ought to have used a different inference rule.

This brings us to our second question: what, beyond keeping an open mind, was the executive required to do instead? When we turn from belief to inquiry, we can see that there are many things the executive could do. He could attend to the totality of perceptually available evidence, looking for cues which suggest that the employee may not be a secretary. He could gather more evidence before making a judgment. He could retrieve relevant information from memory, trying to recall whether he had ever interacted with this woman before. He could process the available evidence through a more demanding inference rule, instead of relying on the simple schema which judges female-presenting employees to be secretaries. And he could take steps during his interactions with female employees to increase his likelihood of remembering information about their professional status so he will not have to rely on stereotypes. The RRCV can explain why many of these steps may be required despite their high cognitive cost, in order to promote a society free from bias and discrimination.

In Sections 7-8, we have seen how the RRCV meets the minimal criteria. The RRCV is stakes- and tradeoff-sensitive, building stakes and tradeoffs directly into the injunction to promote value. This allows the RRCV to make a range of novel and plausible normative predictions, such as the conditions under which an anchoring bias may result from rational inquiry. The RRCV also explains the irrationality of many cases of stereotyping. Stereotyping is often irrational because it imposes significant epistemic and non-epistemic harms throughout all walks of life. And the RRCV explains what we may be rationally required to do instead: not only to bear in mind that individuals may be exceptions to our stereotypes, but also to actively gather and attend to evidence; retrieve information from memory; replace stereotypes with more cognitively demanding forms of inference; and to seek out individuating information that will allow us to avoid stereotyping in the future. Insofar as these results are plausible, they lend support to a reason-responsiveness,

consequentialist view of the rationality of inquiry for bounded agents.

References

Adams, Fred. 2018. "Cognition wars." *Studies in History and Philosophy of Science Part A* 68:20–30.

Ahlstrom-Vij, Kristoffer and Dunn, Jeffrey. 2014. "A defence of epistemic consequentialism." *Philosophical Quarterly* 64:541–551.

Arkes, Hal, Gigerenzer, Gerd, and Hertwig, Ralph. 2016. "How bad is incoherence?" *Decision* 3:20–39.

Arpaly, Nomy and Schroeder, Timothy. 2015. "A causal theory of acting for reasons." *American Philosophical Quarterly* 52:103–14.

Basu, Rima. 2019. "Radical moral encroachment: The moral stakes of racist beliefs." *Philosophical Issues* 29:9–23.

Bermúdez, José. 2009. *Decision theory and rationality*. Oxford University Press.

Bernheim, B. Douglas, Ray, Debray, and Yeltekin, Sevin. 2015. "Poverty and self-control." *Econometrica* 83:1877–1911.

Bertrand, Marianne and Mullainathan, Sendhil. 2004. "Are Emily and Greg more employale than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review* 94:991–1013.

Bodenhausen, Galen. 1990. "Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination." *Psychological Science* 1:319–22.

Bodenhausen, Galen and Lichtenstein, Meryl. 1987. "Social stereotypes and information-processing strategies: The impact of task complexity." *Journal of Personality and Social Psychology* 52:871–80.

Bolinger, Renée. forthcoming. "The rational impermissibility of accepting (some) racial generalizations." *Synthese* forthcoming.

Bradley, Ben. 2006. "Against satisficing consequentialism." *Utilitas* 18:97–108.

Bratman, Michael. 1987. *Intentions, plans, and practical reasoning*. Harvard University Press.

Byron, Michael. 1998. "Satisficing and optimality." *Ethics* 109:67–93.

Callaway, Frederick, Lieder, Falk, Das, Priyam, Gul, Syan, Krueger, Paul M., and Griffiths, Thomas L. 2018. "A resource-rational analysis of human planning." In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, volume 1, 178–183.

Cariani, Fabrizio. 2016. "Deontic modals and probabilities: One theory to rule them all?" In Nate Charlow and Matthew Chrisman (eds.), *Deontic modality*, 11–46. Oxford University Press.

Cherniak, Christopher. 1981. "Minimal rationality." *Mind* 90:161–183.

—. 1986. *Minimal rationality*. MIT Press.

Cohen, Sheldon, Doyle, William, and Baum, Andrew. 2006. "Socioeconomic status is associated with stress hormones." *Psychosomatic Medicine* 68:414–420.

Deck, Cary and Jahedi, Salar. 2015. "The effect of cognitive load on economic decision making: A survey and new experiments." *European Economic Review* 78C:97–119.

Driver, Julia. 2012. *Consequentialism*. Routledge.

Epley, Nicholas and Gilovich, Thomas. 2006. "The anchoring-and-adjustment heuristic: Why the adjustments are insufficient." *Psychological Science* 17:311–318.

Evans, Gary and Schamberg, Michelle. 2009. "Childhood poverty, chronic stress, and adult working memory." *Proceedings of the National Academy of Sciences* 106:6545–6549.

Friedman, Jane. 2019. "Teleological epistemology." *Philosophical Studies* 176:673–691.

—. forthcoming a. "The epistemic and the zetetic." *Philosophical Review* forthcoming.

—. forthcoming b. "Zetetic epistemology." In Baron Reed and A.K. Flowerree (eds.), *Towards an expansive epistemology: Norms, action, and the social sphere*. Routledge.

Gailliot, Matthew, Baumeister, Roy C., Dewall, Nathan, Maner, Jon, Plant, E. Ashby, Tice, Dianne, Brewer, Lauren, and Scheichel, Brandon. 2007. "Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor,." *Journal of Personality and Social Psychology* 92:325–336.

Gardinier, Georgi. 2018. "Evidentialism and moral encroachment." In Keith McCain (ed.), *Believing in accordance with the evidence*, 169–195. Springer.

Gendler, Tamar Szabó. 2011. "On the epistemic costs of implicit bias." *Philosophical Studies* 156:33–63.

Gigerenzer, Gerd. forthcoming. "Axiomatic rationality and ecological rationality." *Synthese* forthcoming.

Gigerenzer, Gerd and Brighton, Henry. 2009. "Homo heuristicus: Why biased minds make better inferences." *Topics in Cognitive Science* 1:107–143.

Gigerenzer, Gerd and Gaissmaier, Wolfgang. 2011. "Heuristic decision making." *Annual Review of Psychology* 62:451–482.

Gigerenzer, Gerd and Goldstein, Daniel. 1996. "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review* 103:650–669.

Gilbert, Daniel and Hixton, Gregory. 1991. "The trouble of thinking: Activation and application of stereotypic beliefs." *Journal of Personality and Social Psychology* 60:509–517.

Gneiting, Tilmann and Raftery, Adrian. 2007. "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102:359–378.

Goldberg, Sanford. 2016. "On the epistemic significance of evidence you should have had." *Episteme* 13:449–470.

Harman, Gilbert. 1986. *Change in view*. MIT Press.

Hastings, W.K. 1970. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika* 57:97–109.

Hoek, Daniel. forthcoming. "Minimal rationality and the web of questions." In Dirk Kindermann, Peter van Elswyk, and Andy Egan (eds.), *Unstructured content*. Oxford University Press.

Icard, Thomas. 2018. "Bayes, bounds, and rational analysis." *Philosophy of Science* 85:79–101.

Johnson, Eric and Payne, John. 1985. "Effort and accuracy in choice." *Management Science* 31:395–414.

Kagan, Shelly. 2000. "Evaluative focal points." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 134–55. Edinburgh University Press.

Kaplan, Robert, Bhalodkar, Narendra, Brown, Edward, White, Jessica, and Brown, David. 2004. "Race, ethnicity, and sociocultural characteristics predict noncompliance with lipid-lowering medications." *Preventive Medicine* 39:1249–1255.

Kermati, Mehdi, Smittenaar, Peter, Dolan, Raymond, and Dayan, Peter. 2016. "Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum." *Proceedings of the National Academy of Sciences* 113:12868–12873.

Kiesewetter, Benjamin. 2017. *The normativity of rationality*. Oxford University Press.

Kolodny, Niko. 2005. "Why be rational?" *Mind* 114:509–563.

Kolodny, Niko and MacFarlane, John. 2010. "Ifs and oughts." *Journal of Philosophy* 107:115–143.

Korsgaard, Christine. 1997. "The normativity of instrumental reason." In Garrett Cullity and Berys Gaut (eds.), *Ethics and practical reason*, 215–254. Oxford University Press.

Lavin, Douglas. 2004. "Practical reason and the possibility of error." *Ethics* 114:424–457.

Lever, Annabelle. 2005. "Why racial profiling is hard to justify: A response to Risse and Zeckhauser." *Philosophy and Public Affairs* 33:94–110.

—. 2007. "What's wrong with racial profiling? Another look at the problem." *Criminal Justice Ethics* 26:20–28.

Lieder, Falk and Griffiths, Thomas. 2017. "Strategy selection as rational metareasoning." *Psychological Review* 124:762–794.

Lieder, Falk, Griffiths, Thomas, Huys, Quentin, and Goodman, Noah. 2018. "The anchoring bias reflects rational use of cognitive resources." *Psychonomic Bulletin and Review* 25:322–349.

Linde, Lena and Bergström, Monica. 1992. "The effect of one night without sleep on problem-solving and immediate recall." *Psychological Research* 54:127–137.

Lord, Errol. 2017. "What you're rationally required to do and what you ought to do (are the same thing!)." *Mind* 126:1109–1154.

—. 2018. *The importance of being rational*. Oxford University Press.

Mani, Anandi, Mullainathan, Sendhil, Shafir, Eldar, and Zhao, Jiaying. 2013. "Poverty impedes cognitive function." *Science* 341:976–980.

Marewski, Julian and Schooler, Lael. 2011. "Cognitive niches: An ecological model of strategy selection." *Psychological Review* 118:393–437.

Metropolis, Nicholas, Rosenbluth, Arianna W., Rosenbluth, Marshall N., Teller, Augusta H., and Teller, Edward. 1953. "Equation of state calculations by fast computing machines." *Journal of Chemical Physics* 21:1087–1092.

Morton, Jennifer. 2017. "Reasoning under scarcity." *Australasian Journal of Philosophy* 95:543–559.

Moss, Sarah. 2018. *Probabilistic knowledge.* Oxford University Press.

Mulgan, Tim. 2001. "How satisficers get away with murder." *International Journal of Philosophical Studies* 9:41–46.

Munnell, Alicia H., Tootell, Geoffrey M.B., Browne, Lynne E., and McEneaney, James. 1996. "Mortgage lending in Boston: Interpreting HMDA data." *American Economic Review* 86:25–53.

Neuberg, Steven and Fiske, Susan. 1987. "Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes." *Journal of Personality and Social Psychology* 53:431–444.

Olsson, Erik J. and Westlund, David. 2006. "On the role of the research agenda in epistemic change." *Erkenntnis* 65:165–183.

Pager, Devah and Shepherd, Hana. 2008. "Racial discrimination in employment, housing, credit, and consumer markets." *Annual Reviews of Sociology* 34:181–209.

Parfit, Derek. 1984. *Reasons and persons.* Oxford University Press.

Pascoe, Elizabeth A. and Richman, Laura Smart. 2009. "Perceived discrimination and health: A meta-analytic review." *Psychological Bulletin* 135:531–554.

Patel, Nirav, Grandner, Michael, Xie, Dawei, Branas, Charles, and Gooneratne, Nalaka. 2010. "'Sleep disparity' in the population: Poor sleep quality is strongly associated with poverty and ethnicity." *BMC Public Health* 10:e1–11.

Pettit, Becky and Western, Bruce. 2004. "Mass imprisonment and the life course: Race and class inequality in U.S. incarceration." *American Sociological Review* 69:151–169.

Pettit, Philip. 1984. "Satisficing consequentialism." *Proceedings of the Aristotelian Society, Supplement* 58:165–176.

Pettit, Philip and Smith, Michael. 2000. "Global consequentialism." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 121–133. Edinburgh University Press.

Pierson, Emma, Simoiu, Camelia, Overgoor, Jan, Corbett-Davies, Sam, Jenson, Daniel, Shoemaker, Amy, Ramachandran, Vignesh, Barghouty, Phoebe, Phillips, Cheryl, Shroff, Rvavi, and Goel, Sharad. forthcoming. "A large-scale analysis of racial disparities in police stops across the United States." *Nature Human Behavior* forthcoming.

Portmore, Douglas. 2011. *Commonsense consequentialism: Wherein morality meets rationality*. Oxford University Press.

Railton, Peter. 1984. "Alienation, consequentialism, and the demands of morality." *Philosophy and Public Affairs* 13:134–171.

Rinard, Susanna. 2019. "Equal treatment for belief." *Philosophical Studies* 176:1923–1950.

Risse, Mathias. 2007. "Racial profiling: A reply to two critics." *Criminal Justice Ethics* 26:4–19.

Risse, Mathias and Zeckhauser, Richard. 2004. "Racial profiling." *Philosophy and Public Affairs* 32:131–170.

Schmidt, Andreas. 2019. "Getting real on rationality – behavioral science, nudging, and public policy." *Ethics* 129:511–543.

Schmitt, Michael T., Branscombe, Nyla R., Postmes, Tom, and Garcia, Amber. 2014. "The consequences of perceived discrimination for psychological well-being: A meta-analytic review." *Psychological Bulletin* 140:921–948.

Schroeder, Mark. 2018. "Can beliefs be wrong?" *Philosophical Topics* 46:115–127.

Serwe, Sascha and Frings, Christian. 2006. "Who will win Wimbledon? The recognition heuristic in predicting sports events." *Behavioral Decision Making* 19:321–332.

Shah, Anuj, Mullainathan, Sendhil, and Shafir, Eldar. 2012. "Some consequences of having too little." *Science* 338:682–685.

Shenhav, Amitai, Musslick, Sebastian, Lieder, Falk, Kool, Wouter, Griffiths, Thomas, Cohen, Jonathan, and Botvinick, Matthew. 2017. "Toward a rational and mechanistic account of mental effort." *Annual Review of Neuroscience* 40:99–124.

Simmons, Joseph, LeBoeuf, Robyn, and Nelson, Leif. 2010. "The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors?" *Journal of Personality and Social Psychology* 99:917–932.

Simon, Herbert. 1976. "From substantive to procedural rationality." In T.J. Kastelein, S.K. Kulpers, W.A. Nijenhuls, and R.G. Wagenaar (eds.), *25 years of economic theory*, 65–86. Springer.

Sims, Christopher. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50:665–690.

Slote, Michael. 1984. "Satisficing consequentialism." *Proceedings of the Aristotelian Society, Supplement* 58:139–163.

Staffel, Julia. 2020. *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.

Steffensmeier, Darrell and Demuth, Stephen. 2000. "Ethnicity and sentencing outcomes in U.S. federal courts: Who is punished more harshly?" *American Sociological Review* 65:705–729.

Stein, Edward. 1996. *Without good reason: The rationality debate in philosophy and cognitive science*. Clarendon Press.

Stich, Stephen. 1990. *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation.* MIT Press.

Stigler, George. 1961. "The economics of information." *Journal of Political Economy* 69:213–225.

Sylvan, Kurt. 2020. "An epistemic non-consequentialism." *Philosophical Review* 129:1–51.

Thomsen, Frej Klem. 2011. "The art of the unseen: Three challenges for racial profiling." *Journal of Ethics* 15:89–117.

Todd, Peter, Fiddick, Laurence, and Krauss, Stefan. 2000. "Ecological rationality and its contents." *Thinking and Reasoning* 6:375–384.

Troffaes, Matthias C.M. 2007. "Decision making under uncertainty using imprecise probabilities." *International Journal of Approximate Reasoning* 45:17–29.

Tversky, Amos and Kahneman, Daniel. 1974. "Judgment under uncertainty: Heuristics and biases." *Science* 185:1124–1131.

Vohs, Kathleen. 2013. "The poor's poor mental power." *Science* 341:969–970.

Vul, Edward, Goodman, Noah, Griffiths, Thomas L., and Tenenbaum, Joshua B. 2014. "One and done? Optimal decisions from very few samples." *Cognitive Science* 38:599–637.

Wheeler, Gregory. forthcoming. "Less is more for Bayesians, too." In Riccardo Viale (ed.), *Routledge handbook on bounded rationality*. Routledge.

Zack, Naomi. 2015. *White privilege and black rights: The injustice of US police racial profiling and homicide.* Rowman and Littlefield.

Zilberstein, Shlomo, Washington, Richard, Bernstein, Daniel S., and Mouddib, Abdel-Illah. 2001. "Decision-theoretic control of planetary rovers." In Michael Beetz, Joachim

Hertzberg, Malik Ghallab, and Martha E. Pollack (eds.), *Advances in plan-based control of robotic agents*, 270–289. Springer.