inquiry *under* bounds

DAVID THORSTAD

OXFORD

Inquiry under bounds

Inquiry under bounds

DAVID THORSTAD

To Leo, the most rational being I know

Contents

PART IV. APPLYING THE ACCOUNT

Introduction

> Human rational behavior ... is shaped by a scissors whose two blades
> are the structure of task environments and the computational capabil-
> ities of the actor.
>
> <div align="right">Herbert Simon (1990, p. 7)</div>

1 Two turns in the study of human rationality

Humans are twice-bounded creatures. First, we are bounded from within. We have
limited physical and mental abilities, and in exercising our abilities we incur costs.
Second, we are bounded from without. We are born into environments that are not
of our choosing. These environments determine the mental and physical problems
that we face throughout our lives and shape the reward structures of strategies we
bring to bear on those problems.

It is universally acknowledged that theories of rationality should incorporate
our most important physical bounds. Suppose you are planning your route to
work. What it is rational for you to do is limited by your abilities. You cannot be
rationally required to fly to work because you do not have wings. Rational action
is also limited by the costs of physical movement. It would be irrational to run
to work because you would arrive exhausted and sweaty. And rational action is
limited by your environment. You must look both ways before crossing the street
because the street is populated with cars.

The tradition of bounded rationality holds that theories of rationality should
also incorporate our most important cognitive bounds. Suppose you are bagging
groceries. How it is rational for you to cognize is limited by your abilities. You
cannot be rationally required to calculate the optimal arrangement of groceries,
because only an extremely precocious mathematician could do this. Rational
cognition is also limited by the costs of cognition. Unless you are unusually
fond of mathematics, it would be wastefully irrational to spend your time in the
grocery line thinking about how to bag groceries, because your time and attention
could be better spent planning dinner. And rational cognition is limited by your
environment. The reason why it is irrational to plan your bagging strategy is that
there are plenty of bags to go around.

My goal in this book is to develop the consequences of two turns in the study of human rationality. The first of these turns is the *bounded turn*. Because humans are bounded agents, the bounded turn holds that theories of rationality should say explicitly and in detail how our most important bounds bear on rational cognition. These include not only internal bounds such as limited abilities and cognitive costs, but also external bounds such as the structure of the environment. The bounded turn is traditionally coupled with a second *procedural turn* within the study of human rationality.

Herbert Simon held that the fundamental turn in the study of bounded rationality is the turn from *substantive* to *procedural rationality* (Simon 1976). Theories of substantive rationality ask normative questions about the attitudes that result from inquiry. For example, we might ask what it is rational for agents to intend, prefer, or believe. Theories of procedural rationality take a step back, asking normative questions directly about the processes of inquiry through which our attitudes are produced. For example, we might ask when it is rational for agents to gather evidence before making up their minds, or when it is rational to use frugal heuristics to simplify reasoning.

Simon held that many of our most important cognitive bounds, such as limited cognitive abilities and deliberation costs, are most strongly felt as bounds on processes of inquiry rather than as bounds on the attitudes that result from inquiry. As a result, Simon urged theories of bounded rationality to take a procedural turn toward the study of rational inquiry. If Simon was right, then the centerpiece of a theory of human rationality is an account of inquiry under bounds. Such a theory would say how bounded agents are rationally required to inquire. In this book, I motivate, develop, and apply an account of inquiry under bounds.

2 The way forward

My discussion has four parts. Part 1, *Rationality at the Crossroads*, develops the bounded and procedural turns by contrasting two general approaches to the study of rationality: bounded rationality and the Standard Picture.

Chapter 1, *The Standard Picture* introduces the Standard Picture of rationality dominant in many areas of social science and formal philosophy. On the Standard Picture, rationality is captured by requirements of consistency exemplified by logic, probability theory, and decision theory. Empirical findings suggest that humans regularly violate Standard Picture requirements in behavior and cognition. We can react to these findings in two ways. We can blame the agent, retaining the Standard Picture as a normative theory and holding that humans often think and act irrationally. Or we can blame the theory and hold that many observed Standard Picture violations are not straightforward irrationalities, but

rather signs that the Standard Picture is incorrect or incomplete. I suggest that in many cases we should blame the theory.

Chapter 2, *Bounded Rationality* presents an alternative approach. I characterize the bounded tradition using a collection of five claims. First, bounds matter: theories of rationality should say explicitly and in detail how our most important bounds bear on rational cognition. Second, theories of rationality should be process-focused, concentrating on norms governing processes of inquiry rather than the judgments and decisions that result. Third, rational inquiry is often heuristic, using only a subset of available information and processing that information frugally. Fourth, the rationality of cognition is ecological, or environment-relative. And fifth, the right account of rationality should vindicate the rationality of many Standard Picture violations.

Theories of bounded rationality encounter a challenge. Whereas the Standard Picture presents a variety of complete, or nearly complete theories of rationality, the bounded tradition gives us only a collection of normative claims and a tradition of applying them. We need a theory of rational inquiry in order to ground, clarify, and apply these claims, as well as to assess the relationship between bounded rationality and the Standard Picture. I take up this project in Part 2, *Norms of Inquiry*.

Chapter 3, *An Account of Rational Inquiry* develops a reason-responsive consequentialist view (RRCV) of rational inquiry as a special case of a more general global consequentialist program. As a theory of rightness, global consequentialism holds that we ought to do what is best and that what is best is to promote as much value as we can. Global consequentialism has been defended not only as an account of rightness but also as an account of many other normative statuses such as rationality and virtue. I develop a reason-responsiveness conception of rationality and use the reason-responsiveness conception to extend global consequentialism into an account of rational agency. As a special case of this account, we recover a consequentialist, reason-responsive account of rational inquiry.

An immediate objection to this account is that it answers the wrong question. The reason-responsive consequentialist view is an account of rationality simpliciter, not epistemic rationality. Chapter 4, *There Are No Epistemic Norms of Inquiry* responds to this objection. I argue that the same grounds which led us to posit a distinction between epistemic and non-epistemic norms of belief should make us skeptical of the existence of a distinction between epistemic and non-epistemic norms of inquiry. In the case of belief, we may have good grounds to hold that all reasons for belief are epistemic; that ordinary language tracks a distinct type of epistemic rationality; or that the notion of epistemic rationality is needed to play key theoretical roles associated with the concept of rationality. But none of these arguments generalize to motivate the existence of a distinctively epistemic type of rational inquiry, and in fact, I argue, several of these arguments

tell against the existence of epistemic norms governing inquiry. I conclude by sketching an alternative Gibbardian picture of inquiry on which rational inquiry is an all-things-considered affair.

The task of Part 2 was to develop my account of rational inquiry, the reason-responsive consequentialist view. Part 3, *Justifying the Account* gives two arguments for the RRCV.

Chapter 5, *The Argument from Minimal Criteria*, argues that the reason-responsive consequentialist view is our best hope for meeting three minimal criteria on an account of boundedly rational inquiry: sensitivity to tradeoffs made during inquiry; sensitivity to the stakes of inquiry; and explaining the rational impermissibility of many inferences made by stereotyping. I review three existing accounts of boundedly rational inquiry: the Standard Picture, pragmatism, and an account on which inquiry aims at knowledge. I argue that these accounts struggle to satisfy the minimal criteria, while by contrast, the RRCV performs well by the lights of all three criteria.

Chapter 6, *The Explanatory Argument* presents a programmatic explanatory argument for the reason-responsive consequentialist view. Historically, consequentialists have defended their view on the basis of its ability to recover clear, correct, and unifying explanations of a diverse array of normative phenomena. The explanatory argument shows how the RRCV can account for a range of normative claims about inquiry that have eluded philosophical analysis including norms of friendship; duties to avoid cognitive clutter; and norms of logical non-omniscience. In traditional consequentialist fashion, the explanatory argument takes these explanatory successes as evidence for the correctness of the RRCV.

Having developed and defended an account of rational inquiry, Part 4, *Applying the Account*, sets out to reap what we have sown. The first order of business is to use the reason-responsive consequentialist view to articulate and defend the normative claims about bounded rationality set out in Chapter 2. I tackle this project in two stages.

Chapter 7, *Vindicatory Epistemology*, takes up the vindicatory project of showing how the right theory of bounded rationality can be used to vindicate the rationality of many observed Standard Picture violations. I set out four desiderata on a vindicatory program and three strategies that can be used to meet those desiderata. Then I use those strategies together with the reason-responsive consequentialist view to show how two types of Standard Picture violations can be vindicated: biases resulting from a heuristic known as anchoring and adjustment, and errors in conditional reasoning. I use these case studies to illustrate how the reason-responsive consequentialist view grounds a robust program of vindicatory epistemology.

Chapter 8, *Bounded Rationality Revisited*, takes up the remaining four normative claims from Chapter 2. I show how the reason-responsive consequentialist view makes space for the claim that bounds matter by telling a principled

story about how and why bounds matter to rational cognition. I use the global consequentialist commitments of the RRCV to interpret and defend the claim that rationality should be process-focused. I show how the RRCV grounds the rationality of paradigmatic cognitive heuristics and sheds light on the factors guiding rational choice among heuristic strategies. And I use the RRCV to defend an interpretation of ecological rationality.

Chapter 9, *The Standard Picture Revisited*, takes a new look at the relationship between bounded rationality and the Standard Picture. I show how the reason-responsive consequentialist view charts a third way between the two reactions to Standard Picture violations outlined in Chapter 1. This third way leaves the Standard Picture largely intact as a theory of rational attitudes while rejecting the Standard Picture as a guide to rational inquiry. A natural objection is that this approach is too concessive to the Standard Picture because it makes cognitive bounds less relevant to the rationality of attitudes. I argue that my concessive approach is needed to avoid two revisionary consequences: that most traditional attitudinal norms are false, and that all attitudinal norms are normatively non-fundamental.

Chapter 10, *The Zetetic Turns*, relates my project to a recent *zetetic turn* in epistemology from the study of doxastic attitudes toward the study of theoretical inquiry. I argue that the zetetic turn in epistemology should be seen as part of a broader procedural turn in the study of bounded rationality, from the study of attitudes toward the study of inquiry. If this is right, then we can draw lessons for the epistemology of inquiry by reflecting on what we have learned about bounded rationality. I show how the discussion in this book yields five insights for the epistemology of inquiry. I close by noting that our practical attitudes, no less than our doxastic attitudes, are produced by bounded and constrained processes of inquiry. Taking the procedural turn requires making a second zetetic turn within practical philosophy, from the study of practical attitudes toward the study of practical inquiry.

PART I

RATIONALITY AT THE CROSSROADS

1

The Standard Picture

1.1 The emergence of the Standard Picture

The middle of the twentieth century was characterized by profound optimism about human rationality. On the emerging Standard Picture (Stein 1996) of rationality, rational requirements were exemplified by requirements of consistency derived from logic, probability theory, and decision theory. These requirements were at once proposed as descriptive and normative accounts of human behavior and cognition. On the Standard Picture, rationality requires us to think and act consistently, and descriptive theorists held that this is by and large what we do.

Decision theory came to prominence as a descriptive foundation for the newly mathematicized field of economics. This growth was driven by the twin tools of revealed preference theory and representation theorems. The revealed preference approach of Paul Samuelson (1938) and Hendrik Houthakker (1950) showed how the fundamental decision-theoretic concept of preference is revealed by agents' choices under the assumption that agents choose rationally. A series of representation theorems showed in turn how conditions on rational preference could be reformulated as conditions on rational credence and subjective utility (Bolker 1966; Jeffrey 1978; Savage 1954; Von Neumann and Morgenstern 1947). These developments gave rise to a rationality-centered approach to the explanation of behavior. Through revealed preference theory, agents' preferences could be inferred on the basis of their choices, given rationality assumptions. Representation theorems would then redescribe revealed preferences in the language of credence and utility, and that representation could be used to predict future behavior. This three-way relationship between choice, preference, and credence and utilities, mediated by rationality assumptions, enabled economists to predict an increasingly large range of phenomena with high precision and accuracy.

Some theorists cautioned against mistaking these new decision-theoretic assumptions for normative claims. Milton Friedman argued that:

> Positive [descriptive] economics is in principle independent of any particular ethical position or normative judgments. As Keynes says, it deals with "what is," not with "what ought to be." Its task is to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances.   (Friedman 1953, p. 4)

Nevertheless, it became increasingly tempting to take the axioms of rational preference as normative claims. For example, Leonard Savage prefaced his representation theorem by urging the reader to accept the proposed axioms as normative conditions on rational preference.

> I am about to build up a highly idealized theory of the behavior of a "rational" person with respect to decisions. In doing so I will, of course, have to ask you to agree with me that such and such maxims of behavior are "rational"... When certain maxims are presented for your consideration, you must ask yourself whether you try to behave in accordance with them, or, to put it differently, how you would react if you noticed yourself violating them.   (Savage 1972, p. 7)

Here Savage argues that we can recognize the preference axioms as normative requirements by reflecting on the fact that when we notice ourselves violating the preference axioms, we feel that we should revise our preferences to bring them in line with the axioms.

Mid-century decision theory absorbed and popularized the probability calculus as a theory of fine-grained belief. Around the same time, the newly rigorized theory of logic gained popularity as a theory of coarse-grained belief.[1] As Piaget revived the study of cognitive psychology, he gave pride of place to the development of logical reasoning in his theory of human cognitive development (Piaget 1957). From there, logic became integral to the psychology of reasoning, featuring prominently in leading theories (Rips 1994) and tasks (Chapman and Chapman 1959; Wason 1968).

Some theorists resisted this descriptive construal of logic. For example, Frege vehemently criticized what he saw as the psychologism of his contemporaries who understood logic as a descriptive theory of human thought. But Frege readily granted that logic is normative: the laws of logic "prescribe universally the way in which one ought to think if one is to think at all" (Frege 1893/1903, p. xv).[2]

The emergence of the Standard Picture was no accident. There were, and continue to be strong motivations for adopting the Standard Picture as a normative and descriptive account of rationality. My focus in this book is on the Standard Picture as a normative theory, however, it will be important to understand the normative development of the Standard Picture in the light of its descriptive applications and emerging challenges.

Section 1.2 reviews arguments for the Standard Picture as a normative theory. Section 1.3 presents a series of descriptive challenges to the Standard Picture.

---

[1] By the mid-1930s, modern versions of natural deduction had been introduced (Gentzen 1935) and standard metamathematical results such as soundness, completeness (Gödel 1929), and compactness (Gödel 1930) were in place.
[2] Frege's claim survives to this day in the claim that the laws of logic are constitutive norms for thought (Leech 2015; MacFarlane 2002).

We can respond to these challenges in two ways: by retaining the Standard Picture and seeing the descriptive challenges as instances of irrational cognition (Section 1.4), or by using the descriptive challenges as a guide to identify normatively relevant factors to which the Standard Picture is insensitive (Section 1.5).[3]

1.2 Arguments for the Standard Picture

The popularity of the Standard Picture as a normative theory is no accident. There are at least four strong arguments that can be given in favor of the Standard Picture.

A traditional argument for the Standard Picture begins with Dutch Books and money pumps. Agents whose preferences violate many Standard Picture assumptions will prefer to make each of a series of bets which together constitute a Dutch Book (Adams 1962; Lewis 1999; Ramsey 1926/1931), a series of bets which guarantees a sure loss for the agent.[4] Similarly, agents whose preferences are intransitive will be subject to a money pump: they will prefer to make each of a series of trades which leave them back where they started less some fees paid during the exchange (Davidson et al. 1955).[5] Traditionally, Dutch Books and money pumps were taken to show that agents have practical reason to comply with Standard Picture norms in order to avoid exploitation. More recently, philosophers have argued that Dutch Books and money pumps reveal a type of practical incoherence, interpreted by the Standard Picture as a sign of irrationality, in which agents deem as fair a series of bets which are not fair (Armendt 1993; Christensen 1996). Note well that these arguments are not restricted to unbounded agents: many Dutch Book arguments can be reformulated as arguments for approximate coherence requirements on bounded agents (Schervish et al. 2000, 2002, 2003; Staffel 2015).

Recent accuracy-centered arguments have aimed to show how Standard Picture requirements on belief can be supported by considerations of accuracy, rather than by appeals to pragmatic costs of practical incoherence. These arguments begin with principles relating rationality to accuracy and show how these principles can be used to derive norms of rational credence. Most famously, a simple principle of non-dominance forbidding agents from adopting a credence function when some other credence function is guaranteed to be more accurate in any possible world is sufficient to establish the probability axioms as rational requirements (Joyce 1998, 2009). The program has been extended to account for diachronic coherence constraints such as updating by Bayesian conditionalization (Greaves and Wallace

---

[3] To clarify, I use talk of 'blame' here in the sense of assigning fault, not in the philosophers' sense of assessing blameworthiness. I do not mean to make the implausible claim that theories can be blameworthy. Thanks to a referee for pushing me to clarify this.

[4] On the notion of a sure loss see Briggs (2009) and Mahtani (2012, 2015).

[5] Intransitive preferences are preferences of the form $X \succ Y, Y \succ Z, Z \succ X$.

2006; Leitgeb and Pettigrew 2010a, b). And more generally, the program has been used to recover norms that may with some right be regarded as principles of structural rationality, for example, principles of indifference requiring ignorant agents to treat a range of propositions alike (Leitgeb and Pettigrew 2010a, b; Pettigrew 2016) and chance-credence norms linking beliefs about propositions to beliefs about their chances (Pettigrew 2012, 2013). As before, many of these same arguments can be extended to justify approximate coherence requirements on bounded agents (De Bona and Staffel 2017, 2018; Staffel 2020).

A third argument for the Standard Picture begins with representation theorems (Christensen 2001).[6] These theorems allow rational requirements on belief and desire to be decomposed into axiomatic requirements on preference. We can then support the normativity of Standard Picture requirements by arguing for the normativity of the preference axioms. As we saw, Savage urges his readers to accept his theory in precisely this way: we realize that the axioms are rational requirements by reflecting on the fact that if we noticed ourselves violating the axioms, we would be disposed to revise our views. Savage takes this as good evidence that we already endorse the axioms as rational requirements. More generally, we might be led to accept the axioms because they are intuitively plausible, or because learned experience in decision theory convinces us of the strangeness and irrationality of behaviors that result as we weaken the axioms.

Fourth, the Standard Picture can be judged by its results. Within philosophy, the Standard Picture has underwritten enormously successful normative programs including Bayesian epistemology, confirmation theory and decision theory (Buchak 2016; Weisberg 2015). Outside of philosophy, the Standard Picture has generated plausible normative models throughout almost all domains of academic inquiry. For example, the Standard Picture yields good models of rational evidence-gathering (Howard 1968; Stigler 1961), awareness-growth (Karni and Vierø 2013), credence in conditionals (Bradley 2017), control of cognition (Shenhav et al. 2013), and reaction to risk (Buchak 2013). It would not be an exaggeration to say that the Standard Picture is among the most successful normative traditions in the twentieth century in terms of its ability to generate plausible, detailed and correct normative predictions. Together, these predictions generate a plausibility argument for the Standard Picture.

In this section, we have seen that there are at least four strong normative arguments for the Standard Picture. The Standard Picture is supported by pragmatic and coherence-based arguments from Dutch Books and money pumps; accuracy-based arguments in epistemology; arguments from representation theorems; and by a programmatic explanatory argument drawing on the normative plausibility of Standard Picture models.

---

[6] This argument relies on additional premises (Zynda 2000) which may be controversial. For example, agents who are representable as obeying the Standard Picture will also be representable as violating the Standard Picture (Hampton 1994; Meacham and Weisberg 2011). On what grounds is the Standard Picture representation privileged?

1.3 Descriptive trouble for the Standard Picture

Mid-century optimism presented the Standard Picture as a unified descriptive and normative theory of rationality. As the Standard Picture gained in popularity, pressure mounted to subject the Standard Picture to rigorous empirical test. Beginning in the 1960s, researchers uncovered a range of descriptive anomalies that are difficult to explain using Standard Picture models. It is important to survey these anomalies in some detail in order to get a better sense of what has and has not been shown, as well as to open discussion of the many different normative reactions we can have to each anomalous finding.

In probabilistic reasoning, we neglect base rates (Eddy 1982; Kahneman and Tversky 1973; Nisbett et al. 1976), making inferences about traits and events without due concern for their prevalence in an environment. We commit gambler's fallacies (Tversky and Kahneman 1971) and perhaps also hot-hand fallacies (Gilovich et al. 1985), expecting bad luck to reverse and good luck to continue. We neglect sample size (Tversky and Kahneman 1971), making strong inferences from small samples. We systematically misunderstand fundamental statistical phenomena (Tversky and Kahneman 1982). And we make poor judgments about and in the presence of randomness (Bar-Hillel and Wagenaar 1991).

In decisionmaking, we choose dominated options (Birnbaum 2008; Tversky and Kahneman 1986), neglecting options that are better in every respect than the option chosen. We overweight small probabilities and underweight large probabilities (González and Wu 1999; Kahneman and Tversky 1979). We give significant extra weight to certain over merely probable outcomes (Kahneman and Tversky 1979) and extend this weight to outcomes which are not actually certain but only appear so (Tversky and Kahneman 1981). We incorporate sunk costs into decisionmaking (Arkes and Blumer 1985). And we neglect features that we care about in decisionmaking such as the scope (Desvousges et al. 1992) and duration (Fredrickson and Kahneman 1993) of outcomes.

Humans commit a number of logical errors. For example, we judge a conjunction to be more probable than one of its conjuncts (Tversky and Kahneman 1982, 1983) or a disjunction to be less probable than a disjunct (Bar-Hillel and Neter 1993). We are more likely to judge that an argument is deductively valid if we believe its conclusion (Evans et al. 1983). Our knowledge is not closed under logical deduction. There is some evidence that we commit basic errors in conditional reasoning such as denying the antecedent and affirming the consequent (Schroyens et al. 2001).[7]

Sometimes our reasoning is explicitly nonextensional. Our judgments and decisions are subject to framing effects (Chong and Druckman 2007; Tversky and Kahneman 1981) in which the same situation is assessed differently depending on how it is presented. Assessments are also dependent on the format in which

---

[7] See Oaksford and Chater (2003, 2007) and Elqayam and Over (2013) for pushback.

information is presented. For example, we reason differently with information about frequencies of occurrence rather than decimals or probabilities (Cosmides and Tooby 1996; Gigerenzer and Hoffrage 1995), and commit format-specific errors such as paying insufficient attention to denominators in reasoning with fractions (Reyna and Brainerd 2008). Reasoning is subject to unpacking effects in which we judge a disjunction to be more probable when assessing its disjuncts separately (Tversky and Koehler 1994).

The Standard Picture predicts that agents will learn from experience by updating their beliefs, but there are a number of persistent failures to learn which are difficult to explain using Standard Picture models. We are sometimes overconfident (Cooke 1991; Oskamp 1965) after receiving extensive feedback. And even firms that have been in business for many years prepare plans with budgets and timetables more closely aligned to the best-case scenario than to average costs and completion times (Buehler et al. 1994).

The Standard Picture incorporates a simplified model of memory on which propositions are learned through conditionalization and then forever recalled and fully utilized to inform future judgments and decisions. This simplification fails to predict a number of important phenomena. Our beliefs persist after the evidence for them is discredited (Ross et al. 1975). Judgments and decisions are preferentially influenced by information that is more readily available for recall (Tversky and Kahneman 1973). This availability, in turn, is influenced by factors such as the recency (Bjork and Whitten 1974) of learning and the context (Godden and Baddeley 1975) of learning and retrieval. Some information is forgotten. Memories can provide false information, as in incorrect eyewitness identifications (Lindsay and Johnson 1989). Memory is influenced by features of the learning context such as the position of information in a list (Murdock 1962).

Decisionmakers demonstrate a number of preferences that are difficult to reconcile with their stable evaluative attitudes. We show endowment effects, preferring possessed items to unpossessed items (Kahneman et al. 1990). We also come to like items through mere exposure to them (Zajonc 1968). These preferences conflict with our global evaluative attitudes, which do not endorse mere exposure or possession as reasons for preference.

On the Standard Picture, decisions are influenced by anticipated but not incidental affect. That is, decisionmakers may take into account how they will later feel about their decisions, but they will not be influenced by their emotional state at the time of decision. However, there is a strong influence of incidental affect on decisionmaking (George and Dane 2016; Slovic et al. 2007). For example, angry decisionmakers become increasingly risk-seeking, perceive risks to be lessened, and perceive negative events to be predictable and caused by others, while by contrast, fearful decisionmakers are less risk-seeking, perceive risks to be heightened and perceive negative events to be unpredictable and under situational control (George and Dane 2016). Other influences of affect on decisionmaking

are likewise not easily accommodated. For example, research on risk and affect reveals a three-way pathway in which positive feelings toward an option at once increase perceived benefits and decrease perceived risks (Fischhoff et al. 1978; Slovic et al. 2007). This finding is difficult to reconcile with the Standard Picture, for example, because agents do not hold the Panglossian belief that benefits and risks are negatively correlated.

This array of Standard Picture violations raises two questions. The first question is what changes, if any, should be made to descriptive theories of rationality to accommodate the violations? I will not be concerned with this descriptive question here, although much of what I have to say can and has been reflected in descriptive theorizing (Gigerenzer and Selten 2001; Simon 1959; Todd and Gigerenzer 2012a). The second question is what changes, if any, should be made to normative theories of rationality to accommodate observed Standard Picture violations? This question will occupy the remainder of the book.

1.4 Retaining the Standard Picture

When faced with descriptive violations of the Standard Picture, we can react in one of two ways. On the one hand, we can blame the agent. We can retain the Standard Picture as a normative theory and hold that descriptive findings reveal irrational behavior or cognition. On the other hand, we can blame the theory. We can use these descriptive violations as guideposts, seeking an explanation of why humans think and act as we do and, if all goes well, using that explanation to rationalize their behavior.

There is a lot to be said in favor of blaming the agent. In this section, I survey five arguments in favor of blaming agents and retaining the Standard Picture as a normative theory.

First, it should be uncontroversial that some observed Standard Picture violations are irrational. A numerate adult who chooses a 7/100 chance of winning a prize over a 1/10 chance of winning has acted irrationally (Denes-Raj and Epstein 1994). In such cases, standard rationalizing explanations ring hollow. This agent does not need to be charitably reinterpreted so that her thoughts make sense.[8] We have understood her aright. And it is cold comfort to claim that this agent's actions sprung from an underlying reasoning competence.[9] In fact, we have been given strong reasons to doubt that she is competent at reasoning with fractions. Nor should we seek hidden meaning in experimental instructions, on the basis of

---

[8] See (Cohen 1981; Sober 1978) in support of charity-based arguments and (Stein 1996; Stich 1990; Thagard and Nisbett 1983) against.

[9] See (Cohen 1981) in favor of competence-based arguments, and (Stein 1996; Stanovich and West 2000) against.

which the agent has made a good Gricean inference that her chance of winning is greater than 7/100. Sometimes such arguments work, but often they do not.[10] In cases such as this, we should freely admit that the agent has acted irrationally.

Second, when confronted with the fact that their beliefs and preferences violate the Standard Picture, agents often withdraw the offending attitudes (Tversky and Kahneman 1983).[11] For example, if I tell you that you hold a set of cyclical preferences or inconsistent beliefs, then this will probably move you to reflect further on your preferences or beliefs, at least if they concern a subject matter that is important to you. A good explanation for your withdrawal is that you believe it is irrational to hold cyclical preferences or inconsistent beliefs. If Standard Picture violations were rational, it seems that you would be well within your rights to retort that you do not plan to revise your attitudes because there is nothing rationally defective about your preferences or beliefs. Most of us are not disposed to respond in this way, and this is some evidence that we take Standard Picture norms to express genuine rational requirements.

Third, many Standard Picture violations correlate negatively with cognitive ability (Stanovich 1999; Stanovich and West 2000).[12] For example, individuals higher in cognitive ability are more likely to obey logical laws in syllogistic reasoning and to obey the probability calculus in many statistical reasoning tasks. A natural explanation for this phenomenon is that individuals higher in cognitive ability are better able to discern what rationality requires. Then the fact that high-ability cognizers more closely follow a given Standard Picture norm can be taken as evidence that this norm expresses a rational requirement.

Fourth, overzealous attempts to defend human rationality have rightly been accused of taking the Panglossian view that we live in the best of all possible worlds, in the sense that humans rarely or never think and act irrationally (Kahneman 1981; Stanovich and West 2000). This is surprising because in other domains, we are readily disposed to admit that humans think and act irrationally. We suffer from major psychological illnesses and delusions. We accept conspiracy theories, join cults, and doubt established scientific facts. We drink, gamble, overspend, and abuse narcotics, often with life-altering consequences. Advocates of the Standard Picture have rightly wondered why, if we are ready to admit that some people overconfidently believe themselves to be the king of France, we should be so hesitant to accept that people are overconfident about more mundane matters such as the location of France on a map.

---

[10] For some purported successes of this strategy see (Dulany and Hilton 1991; Hertwig and Gigerenzer 1999). For some purported failures see (Tversky and Kahneman 1983).

[11] This does not always happen. However, failures of withdrawal often target the most controversial elements of the Standard Picture, so this may pose no deep problem for the argument (Slovic and Tversky 1974).

[12] Caution is needed in interpreting this argument because a large number of Standard Picture deviations show null or positive correlations with cognitive ability.

Finally, defenders of the Standard Picture should demand a systematic alternative. It is a familiar refrain among decision theorists that it is all too easy to criticize theories developed at the level of clarity, detail, and generality of the Standard Picture. What is lacking in many criticisms is a comparably general, powerful and plausible normative alternative. If we are to replace the Standard Picture, we need not merely to observe that people sometimes violate the Standard Picture and suggest that some of these violations are rational, but also to ground our rationalizing explanations in a new and systematic normative theory.

Summing up, there are at least five things to be said in favor of blaming agents and retaining the Standard Picture. Some Standard Picture violations are clearly irrational. Agents often withdraw attitudes upon learning that they violate Standard Picture norms. High-ability cognizers are often more faithful to Standard Picture norms. Attempts to defend human rationality sometimes border on the Panglossian, and should be supported by a systematic alternative theory. These are strong arguments, and in some cases they are decisive. But I hope to show that there is room for another reaction.

1.5 Modifying the Standard Picture

A second reaction to Standard Picture violations is to blame the theory. By thinking carefully about why agents systematically deviate from Standard Picture requirements, we can discover normative factors which agents are sensitive to, but that the Standard Picture ignores.[13]

One thing to note about the Standard Picture is that it is architecturally neutral. That is, Standard Picture requirements are not derived from, and do not depend on, facts about an agent's cognitive architecture. The Standard Picture requires the same logical, probabilistic, and decision-theoretic consistency of humans, superhumans, toddlers and toads. This creates a surprising asymmetry between physical bounds, which are incorporated into the Standard Picture, and cognitive bounds, which are not.

No one would deny that limitations on physical abilities and the costs of exercising them bear on rationality. We cannot be rationally required to fly to work rather than walk, because we lack the ability to fly: we do not have wings. And we cannot be rationally required to run to work rather than walking, because running

---

[13] The view developed in this book will agree with the Standard Picture in one important way: it accepts the normative importance of expected utility maximization (Chapter 3). I may also have a good deal in common with the Standard Picture in my view of the rationality of attitudes (Chapter 9). However, I come apart from the Standard Picture in my view that rational processes of cognition are often heuristic (Chapter 2) and dependent on cognitive architecture (Chapter 1) and task environments (Chapter 2), do not always aim at coherence or structural rationality (Chapter 5), and are not always helpfully characterized in an axiomatic way (Gigerenzer 2019). I revisit the relationship between bounded rationality and the Standard Picture in Chapter 9.

is costly: we would arrive sweaty and exhausted. But we would be rationally required to fly if we had wings or to run if we could do so without effort. All of these facts are neatly incorporated into Standard Picture models as restrictions on the set of available options and changes to the utility function, respectively.

By contrast, the Standard Picture denies that cognitive abilities and the costs of exercising them bear on rationality. Full logical omniscience is required, even if we lack the ability to deduce complex cognitive truths. And it is no objection to logical omniscience requirements that a truth that is within our cognitive abilities to deduce would be costly to discover, perhaps requiring several years of sustained thought. This asymmetry between the Standard Picture's treatment of cognitive and physical bounds looks like a principled reason to begin pushing back against the Standard Picture.

It is a familiar refrain that incorporating cognitive bounds will require some modification to the Standard Picture. For example, we might require agents to make only those logical inferences that are relatively foreseeable, or beneath a certain complexity. And similarly, we might permit agents to assign credence less than one to some logical truths as a form of epistemic modesty, occasioned by the thought that our logical beliefs have been wrong in the past.

What is less well-appreciated is that a large variety of Standard Picture violations, even quite foreseeable ones, can be brought about by rational cognition in bounded agents. Consider, for example, the conjunction fallacy (Tversky and Kahneman 1982, 1983). This occurs when agents assign greater probability to a conjunctive proposition $p \wedge q$ than to one of its conjuncts, p. The conjunction fallacy is as bald, foreseeable, and stark as a Standard Picture violation can get.[14] Could even such a thing be rational?[15]

To see why it might be, consider the original explanation given by Tversky and Kahneman for how the conjunction fallacy arises.[16] Sometimes we want to judge the probability that an object x belongs to a category M. For example, we might want to judge the probability that a distant tree is a birch. One way we can do this is through the *representativeness heuristic*: judge the likelihood that x is an M by considering the degree to which it resembles a representative M.[17] For example, if

---

[14] For a similarly stark violation, consider Gilbert Harman's (1973) discussion of clutter avoidance, in which rational agents may avoid making simple inferences such as the inference from p to $p \vee q$ in order to reduce cognitive clutter.

[15] Ultimately, I will claim that conjunction fallacies cannot be rational. However, the heuristic inquiries that produced them can be rational. And that is what matters. See the discussion of process focus in Chapters 2, 8, and 9.

[16] To be clear: there are many ways in which conjunction fallacies can be produced. The explanation given here is the most popular explanation for how conjunction fallacies are produced in the cases that I will consider.

[17] The representativeness heuristic was proposed by the heuristics and biases tradition. Like most heuristics in this tradition, representativeness is sometimes criticized by the fast-and-frugal heuristics camp for being imprecise or incompletely specified (Gigerenzer 1996). For the most part, my examples will be drawn from the fast-and-frugal heuristics tradition, but I do think it can be illustrative to think

a tree (x) is light-colored, moderately sized and has thin, lenticled bark, then it is probably a birch (M).

But the representativeness heuristic has a problem. The logic of representativeness comes apart from the calculus of probability in startling ways. Most famously, an unrepresentative M may be more representative of the conjunctive category MN than of the simple category M. By contrast, it is an elementary theorem of the probability calculus that nothing can be more likely to be an M than to be an MN. This fact can, on scattered occasions, give rise to a conjunction fallacy.

Suppose I show you personality sketches and ask you to judge the profession and hobbies of the person sketched based only on a brief description. Here it is understandable why you might employ the representativeness heuristic, judging their likely profession by their resemblance to a representative member. That is a cognitively efficient way to make accurate judgments based on limited information. So for example, I might tell you the following about Bill:

> Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.   (Tversky and Kahneman 1982, p. 92)

If you hold typical, if perhaps unfortunate stereotypes about hobbies and professions, you will judge that Bill is an unrepresentative jazz musician (M) but a representative accountant (N). As a result, Bill is more representative of the conjunctive category MN than of the simple category M. At least Bill resembles a representative jazz-playing-accountant in having traits representative of an accountant. If you judge by representativeness, you will then judge that Bill is more likely to be a jazz-playing accountant than to be a jazz player. And that is a conjunction fallacy.

What do we want to say about these cases? On the one hand, they are not happy cases: I will argue in Chapter 3 that we feel some evaluative tension toward agents in these cases. But on the other hand, we should be prepared to admit that rational heuristics such as representativeness, used rationally, can sometimes lead to Standard Picture violations. It is a familiar refrain of process reliabilists that an inference procedure can be rational even if it is known ahead of time to occasionally produce inaccurate results, so long as these inaccuracies are only occasional. So too, we should be willing to consider the possibility that a heuristic inference procedure can be rational even if it is known ahead of time to occasionally produce inconsistent results, so long as these inconsistencies are only occasional. This occasional vulnerability to inconsistency may be compensated by

through ways to extend the reach of bounded rationality into other traditions, even for readers who might wish that heuristics such as representativeness had been specified in another way.

other factors such as accuracy and cognitive efficiency, and in that case, we have the beginnings of a normative case for the rationality of Standard Picture violations.

The claim that heuristics may be rational despite their occasional vulnerability to bias is not so controversial as it may appear. In the paper that launched their heuristics and biases program, Amos Tversky and Daniel Kahneman conceded that it is understandable why agents might make use of frugal heuristics such as representativeness:

> It is not surprising that useful heuristics such as representativeness and availability are retained, even though they occasionally lead to errors in prediction or estimation.   (Tversky and Kahneman 1974, p. 1130)

Three decades later, the editors of a leading anthology on heuristics and biases reiterated this position in no uncertain terms:

> The heuristics themselves are sensible estimation procedures that are by no measure "irrational."   (Gilovich and Griffin 2002, p. 3)

Just one year later, Daniel Kahneman would deliver a Nobel address entitled 'Maps of bounded rationality' in which he cited Herbert Simon as his inspiration and claimed that his life's work, together with Amos Tversky, was aimed at mapping the shape of bounded rationality (Kahneman 2003).

Here, again, we return to the suggestion of taking bounded rationality as a systematic alternative to the Standard Picture. Through careful descriptive examination of the bounds on human cognition together with normative theorizing about the rational impact of those bounds, we can construct a theory that explains how and why agents may rationally deviate from Standard Picture norms in order to achieve valuable cognitive goals. If all goes well, the right theory of bounded rationality will allow us to retell the story of human cognition so that many seeming irrationalities are revealed as the results of boundedly rational cognition.

That is the plan. Here is how we will follow it. Chapter 2 introduces the bounded rationality program, characterizing the program using five normative claims. This discussion will reveal the need for a theory of rational inquiry in order to clarify and defend the bounded rationality program. Parts 2–3 develop (Chapter 3) and defend (Chapters 4–6) a theory of rational inquiry designed to do just that. Part 4 applies the theory of rational inquiry developed in Parts 2–3 to vindicate the rationality of some Standard Picture violations (Chapter 7), ground the remaining claims made by the bounded tradition (Chapter 8), revisit the relationship between bounded rationality and the Standard Picture (Chapter 9), and use what we have learned to illuminate the recent *zetetic turn* within epistemology, as well as to suggest the need for an analogous turn within practical philosophy (Chapter 10).

2

Bounded rationality

2.1 Introduction

Chapter 1 introduced the Standard Picture and raised the possibility that some Standard Picture violations should be blamed not on the agent, but on the theory. In particular, we saw that many Standard Picture violations may be instances of boundedly rational cognition. This suggests that we should look to bounded rationality for an alternative normative theory. And although my focus in this book will be normative rather than descriptive, the same considerations will motivate a change of descriptive theory in many contexts as well.

The basic idea of bounded rationality is clear enough. We have limited resources and abilities, and these limitations bear on how we should think and act. For example, we would not want to decide between running and walking to class in a way that neglects our physical limitations or the cost of physical exercise. So too, we should not select strategies for cognitive tasks such as choosing a breakfast cereal or assessing the truth of a news report without reference to our cognitive abilities and the costs of using them.

In developing the theory of bounded rationality beyond this basic insight, we encounter a problem. The Standard Picture is a complete, or nearly complete normative theory. Many advocates of the Standard Picture take the right set of decision-theoretic preference axioms together with a few additional rationality constraints to capture most or all of what rationality requires. But there is no comparably general account of what it means to be boundedly rational. The bounded tradition gives us not a broad, overarching theory of rationality but rather a collection of normative claims and a tradition of using these claims to inform rationality modeling.

My project in this chapter is to set out five key normative claims characteristically made by the bounded tradition (Thorstad forthcoming c). In stating these claims, we will see the need for a background normative theory in order to unpack the contents of these claims as well as to assess their correctness and their normative implications. Parts 2–3 will develop and defend such a normative theory, and Part 4 will show how that theory bears on the contents, correctness, and implications of these characteristic normative claims. This effort will take us closer to understanding the bounded tradition as a systematic normative alternative to the Standard Picture, and in Part 4 I will reassess the relationship between bounded rationality and the Standard Picture.

2.2 Bounds matter

First and foremost, the bounded tradition holds that *bounds matter*. To say that bounds matter is to place at least two constraints on rationality modeling. The first is a relevance constraint.

> **Relevance constraint:** Paradigmatic cognitive bounds including limitations on agents' cognitive abilities as well as the costs of cognition bear on the rationality of agents' judgment and decisionmaking.

The relevance constraint leaves many questions open for further theorizing. It does not exhaustively list the bounds that bear on rational cognition.[1] And while the relevance constraint provides two examples of relevant bounds, namely cognitive abilities and cognitive costs, it does not say how these bounds are normatively relevant, but only that they are in fact normatively relevant. Further work is needed to answer these and other questions left open by the relevance constraint.

The relevance constraint is opposed to a common view of rational theorizing on which cognitive costs and limited computational abilities are normatively irrelevant. For example, Ralph Wedgwood asks us to understand the credences of a rational agent as follows:

> One picturesque way of conceiving of [the] rational probability function is to imagine an *angel* perched inside the thinker's head—where the angel's advice to the thinker takes the form of this rational probability function. Unfortunately, this angel is uncertain about many empirical propositions about the world. However, the angel knows all relevant truths about the mental states and events that are present in the thinker's mind at the time; and she can assign probabilities to these empirical propositions by relying on what she knows about these mental states and events, together with everything that the essential nature of these mental states and events either guarantees or makes likely to be true.
>
> (Wedgwood 2018, p. 99)

On Wedgwood's view, rationality requires us to believe as we would if our limited mental faculties were replaced by the unlimited mental powers of an angel. Only, most of us do not have angels in our heads. We have limited and costly mental processes which we must use to make good judgments and decisions. The bounds on our mental processes affect how we can and should cognize. As a result, the relevance constraint insists that our bounds should be reflected in rational theorizing.

---

[1] For discussion see Carr (2022).

The claim that bounds matter imposes a second, subtler constraint on rationality modeling. This constraint is methodological rather than normative.

> **Methodological constraint:** Relevant bounds should be incorporated, as far as possible, into all stages of rationality modeling.

The methodological constraint is meant to block models on which we first determine what would be required of unbounded agents, such as Wedgwood's angelically possessed thinker, then adjust the model to determine how bounded agents should cognize.

The methodological constraint is opposed to an early practice in Bayesian theorizing of abstracting away from cognitive bounds until the end of theorizing. For example, here is a canonical statement of the Bayesian six-step method of rational analysis for cognitive modeling.

> (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. (Chater and Oaksford 1999b, p. 59)

The method of rational analysis has much in common with theorizing in the bounded tradition. For example, step two incorporates the claim to be made in Section 2.5: that rationality is ecological, or environment-relative. But rational analysis has been heavily criticized for its third step, which insists that computational limitations should be written out of rationality models as far as possible. If we accept the relevance constraint, then computational limitations bear on rational cognition. Then what principled grounds could there be for excluding tractable specifications of computational limitations from rationality modeling?

Bayesians have increasingly acknowledged this criticism, revising the method of rational analysis to assume from the start that agents respond systematically and in a normatively correct manner to their cognitive bounds. The paradigm of *cognitively bounded rational analysis* (Howes et al. 2009) replaces the third step of rational analysis with an explicit insistence on incorporating relevant features of cognitive architecture into rationality models. More recent Bayesian paradigms such as computational rationality (Gershman et al. 2015), boundedly rational analysis (Icard 2018), and resource-rational analysis (Lieder and Griffiths 2020)

have continued this trend, making substantial normative and descriptive progress by incorporating cognitive bounds at all levels of rationality modeling.

The methodological constraint has made less headway into philosophical discussions of bounded rationality. For example, here is how Julia Staffel describes the methodology of Bayesian rationality modeling:

> Norms are derived by thinking about the characteristic role our credences are supposed to play in our thinking and decision-making ... In developing the norms, *we abstract away from limiting factors that interfere with credences playing this role perfectly, such as processing or time limitations or possibilities of error.* This view of the Bayesian method makes clear why the ideal norms are unreachable for non-ideal thinkers, yet still apply to them: the limiting factors that are abstracted away from in formulating the norms for our attitudes do in fact constrain thinkers like you and me.   (Staffel 2020)

This passage appears to deny the relevance constraint: in claiming that norms that do not incorporate cognitive bounds apply to bounded agents, Staffel might deny the normative relevance of these cognitive bounds. But on a more charitable reading, perhaps Staffel means only to deny the methodological constraint. For Staffel, the right way to do rationality modeling is what the early rational analyst suggests: think first about the function of cognition in a way that abstracts from cognitive bounds, then ask how rational agents should go about approximately fulfilling that function. If the function of cognition is well-characterized by Standard Picture norms, then rational agents should strive to approximate Standard Picture norms as far as possible.[2]

This may seem like an innocuous move, but I hope to convince you that it is quite dangerous. As Bayesian cognitive scientists have increasingly recognized, we will not get a systematic and correct view of how resource- and ability-bounds bear on rational cognition unless those bounds are incorporated at the beginning of theorizing. This point can only be made by example.[3] Throughout this book, I will argue that rational agents often inquire in ways that run a substantial risk of producing judgments and decisions that violate Standard Picture norms, even

---

[2] See also Zynda (1996).

[3] That is not entirely true. There are some general results, such as the general theory of the second best (Daoust 2021; Lipsey and Lancaster 1956; Wiens 2020) which can be cited in favor of the methodological constraint. But a large part of the case for the methodological constraint continues to rest on repeated, domain-specific examples where the approximation of unbounded models leads us normatively astray. The reason for this is that all models must make simplifying assumptions. The methodological constraint claims that a particular sort of simplifying assumption, namely the elimination of paradigmatic cognitive bounds, is often especially misleading, and that it would be better to make other types of simplifying assumptions instead. This claim cannot be derived from general-purpose results about the reliability of approximation. It needs to be motivated by thinking specifically about the reliability of approximation to cognitively unbounded models.

when other methods of inquiry would cleave more closely to the Standard Picture. One part of the argument will be that agents can improve the expected quality of their judgments and decisions by accepting a larger risk of violating some Standard Picture norms (Arkes et al. 2016). If that is right, then we will not get a correct view of bounded rationality by viewing bounded rationality as a matter of approximating Standard Picture norms.

Summing up, the bounded tradition's first contention is that bounds matter. This claim involves a normative constraint, which holds that cognitive bounds are normatively relevant, as well as a methodological constraint, which holds that cognitive bounds should be incorporated as far as possible into all stages of rationality modeling.

2.3 Process focus

Herbert Simon held that a fundamental turn in the study of bounded rationality is the turn 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 the processes of inquiry by which these outcomes are produced.

Substantive and procedural rationality are not new types of normative assessment. Substantive and procedural rationality incorporate familiar types of normative assessment but differ in the objects assessed. At each level, we can ask familiar types of normative questions about the outcomes or process of inquiry. At the substantive level, we can ask for example what a virtuous agent would believe, what it would be fitting for her to desire, what she ought to prefer, or how she is rationally required to act. We can equally well ask the procedural questions of how a virtuous agent would inquire, how it is fitting for her to inquire, how she ought to inquire, or how she is rationally required to inquire.

Most theorists in the bounded tradition have followed Simon in thinking that normative theory should pay central attention to inquiry. One reason for this shift is that resource- and ability-bounds bear most strongly on the process of inquiry itself.[4] There are, to be sure, propositions so complex that we cannot believe them and intentions so revolting that we cannot hold them, but in general the costs of computation and our limited cognitive abilities make themselves manifest during the process of inquiry by which beliefs or intentions are produced. The second claim made by the bounded tradition is therefore that normative

---

[4] For discussion, see Chapter 8.

assessment should be *process focused*, focusing on the process of inquiry rather than its outcomes.[5]

It is important to distinguish two interpretations of the relationship between substantive and procedural rationality. I have been developing an interpretation on which substantive and procedural rationality are complementary. On this interpretation, the distinction between substantive and procedural rationality is a distinction between two classes of objects that we can ask normative questions about: the outcomes of inquiry and the process of inquiry itself. These questions are separate and hence perfectly compatible. On a complementary interpretation, to say that normative assessment should be process focused is to say that we should devote much of our attention in normative theorizing to the process of inquiry rather than its products. Normative questions about outcomes are held to be often misleading, insofar as these questions do not reflect the majority of an agent's bounds, but they are not ill-defined and may be useful for some purposes. It is, for example, no less intelligible or useful to ask whether I ought to buy coffee than to ask how I ought to deliberate about buying coffee.

My complementary interpretation of the distinction between substantive and procedural rationality is quite heretical. Simon drew the distinction between substantive and procedural rationality as follows.

> Behavior is substantively rational when it is appropriate to the achievement of given goals within the limits imposed by given conditions and constraints...
> Behavior is procedurally rational when it is the outcome of appropriate deliberation.   (Simon 1976, pp. 66–7)

Here Simon holds that substantive and procedural rationality assess the same object: the behavior that results from deliberation. Procedural rationality asks one normative question about behavior, namely whether it results from appropriate deliberation. Substantive rationality asks a different normative question about behavior, namely whether it is appropriate when we focus on the first-order decision problem and ignore the process of deliberation. In this way, the notion of procedural rationality substitutes for and replaces the notion of substantive rationality. Most theorists in the bounded tradition have followed what we might call Simon's *substitute* interpretation of procedural and substantive rationality but broadened the scope of assessment from behavior to other mental and non-mental outcomes of decisionmaking.

The distinction between complementary and substitute interpretations is familiar to philosophers as the distinction between direct and indirect normative assessment. The most common forms of indirect normative assessment hold that

---

[5] We will see in Chapter 10 how this claim relates to a recent zetetic turn in the epistemology of inquiry.

the normative status of inquiry determines the normative status of its outcomes. If, for example, I deliberate in a rationally unimpeachable manner about which car to buy, then my resulting intention to buy that car is also rational. By contrast, direct normative assessment allows us to separate normative questions about inquiry and its outcomes. On a direct approach, to say that a process of inquiry is rational is not yet to say anything about the rationality of its outcomes.

One of my goals throughout this book is to advocate what I will call, following Shelly Kagan (2000), an *everywhere direct* approach to normative assessment. I will argue that this everywhere direct approach makes ample room for the Standard Picture at the substantive level, while also revealing the inadequacy of the Standard Picture at the more important procedural level.

2.4 Heuristic rationality

Because bounded rationality is process focused, the bounded tradition concentrates on identifying the processes which rational agents use to make judgments and decisions. Here we arrive at the bounded tradition's third claim:

Heuristic rationality: Agents are often rationally required to inquire heuristically.

Heuristics are frugal strategies that typically do one or both of the following: utilize a small subset of available information or process that information frugally.[6]

The rationality of heuristic inquiry is defended on three grounds. First, agents often face an *accuracy-effort tradeoff*: more demanding forms of inquiry produce better decisions at the cost of increased cognitive and physical effort (Johnson and Payne 1985).[7] Heuristic decisionmaking often strikes the best balance between decision quality and decision cost. Second, agents have *limited abilities*: we cannot always implement demanding Bayesian forms of inquiry no matter how hard

---

[6] It may seem that at this point I have thrown in my hat with a certain approach to theorizing about bounded rationality, namely the fast-and-frugal heuristics tradition of Gerd Gigerenzer and colleagues (Gigerenzer and Selten 2001), to the exclusion of other popular approaches. But in fact, the appeal to heuristic rationality is widely shared. We saw in Chapter 1 that researchers in the heuristics and biases tradition hold that paradigmatic heuristics are often rational. And an increasing number of Bayesians grant that the processes which rational agents use to inquire are often heuristic. For example, Oaksford and Chater (2007) offer a heuristic account of syllogistic reasoning. These theorists use Bayesian tools and other models to shed light on the ways in which agents select heuristics (Lieder and Griffiths 2017), the computations realized by heuristic processes, and other types of rational cognition which are not heuristic. A common motivation for this view is the thought that Bayesian models describe cognition at a relatively high *computational level*, whereas heuristics may describe the lower *mechanistic level* at which Bayesian cognitive processes are implemented (Griffiths et al. 2012).

[7] More generally, we might appeal to a variety of tradeoffs such as speed-accuracy tradeoffs (Heitz 2014) and accuracy-coherence tradeoffs (Thorstad 2021).

we try.[8] Both the possibility and cost of employing demanding inquiry strategies is an agent-relative matter. Decisionmakers with high levels of cognitive ability more often do and should employ Bayesian methods because they are more likely to implement these methods correctly and efficiently (Stanovich 1999; Stanovich and West 2000). Finally, heuristics exploit *less-is-more effects* in which decisions based on a small amount of information often reliably outperform decisions based on all available information (Gigerenzer and Brighton 2009). Very roughly, this is because focusing on a small amount of relevant information avoids overfitting cognitive models to noisy data by focusing on the most important and robust trends in the data (Geman et al. 1992).

When assessing the rationality of heuristic cognition, it is important to ensure that heuristics are fully specified. For ease of exposition, I will discuss the components of decisionmaking heuristics, although throughout this book my emphasis will be at least as much on judgmental heuristics as on decisionmaking heuristics. A fully specified heuristic has three parts (Gigerenzer and Selten 2001b). The first part is a *search rule* which tells agents how to identify options, for example, candidate actions which they could take. The second is a *stopping rule* which tells agents when to halt search. The third is a *decision rule* which says how to choose among options when search halts. For ease of exposition, I will often suppress the search and stopping rules, but they should never be far from mind.

To illustrate the structure of heuristic decisionmaking, suppose you are buying a house. You might make your decision by *satisficing* (Simon 1955). As a search rule, satisficing instructs you to search houses one at a time, perhaps in the order suggested by your realtor. Before beginning search, you specify an *aspiration level* in one or more goods: for example, the house should have at least three bedrooms, two bathrooms, and a garage, and cost no more than four times your current salary. As a stopping rule, satisficing says to halt search when you find a house that meets all of your aspirations. And as a decision rule, satisficing says to buy that house. More complicated versions of satisficing incorporate needed bells and whistles, for example, a procedure of aspiration adaptation in which satisficers adjust their aspirations in light of information retrieved during search (Selten 1998).

The fundamental problem of decisionmaking for bounded agents is strategy selection (Lieder and Griffiths 2017; Marewski and Schooler 2011). In any given decision problem, agents must choose whether to inquire heuristically or non-heuristically, which heuristic to employ, and how to set the internal parameters of the selected heuristic strategy, for example, the initial aspiration level in satisficing.

---

[8] A terminological note: I use talk of limited abilities to track what agents are in-principle capable of doing, to be distinguished from the costs of doing so. This allows us to separate the argument from limited abilities from the accuracy-effort tradeoff, which appeals to the costs of cognition. It is important to separate these arguments, because on my account the accuracy-coherence tradeoff appeals to the principle of expected-utility maximization, whereas the argument from limited abilities relies on the more general claim that ought implies can.

To illustrate the problem of strategy selection, consider three dimensions along which common heuristic strategies vary.

First, heuristics vary in their *compensatoriness*, or the degree to which excesses along one decision cue can make up for shortfalls along another. Satisficing is fully noncompensatory. Our satisficer will not buy a house without a garage, even if it has ten bedrooms and costs a penny. For important decisions such as home-buying, it might be more appropriate to employ a compensatory heuristic that trades off among all modeled decision cues. For example, our home-buyer might decide pairwise between homes by *tallying* the number of decision cues on which each home performs best (Einhorn and Hogarth 1975). In many problems, tallying performs comparably to *linear regression*, which takes a weighted sum of decision cues with weights selected by the method of least squares (Dawes and Corrigan 1974).

Second, heuristics vary in their *cue utilization*. Our satisficer made decisions based on four pieces of information about a house: its cost, number of bedrooms and bathrooms, and the presence of a garage. More frugal heuristics may use only a single decision cue. For example, when choosing between two products or actions you might employ the *recognition heuristic* of choosing the one you have heard of or even the *fluency heuristic* of choosing the option most easily recognized (Goldstein and Gigerenzer 1999). In many contexts, easily recognized options are likely to be reasonably safe and high-quality, justifying this strategy. Similar strategies include the *default heuristic* of taking a habitual or default action without further deliberation, and *take the first* (Johnson and Raab 2003), the strategy of choosing the first option that comes to mind. These heuristics are often appropriate in low-stakes, repeated choice situations such as grocery shopping. They may also be appropriate for experts, as in the chess grandmaster who plays dozens of games at once, often by choosing the first move that comes to mind and trusting her experience to have provided a good move. At the opposite extreme, strategies such as tallying and linear regression may use a large number of decision cues.

Finally, heuristics vary in their *disposition to incoherence* (Thorstad forthcoming). One way to buy a house is through *lexicographic choice* (Fishburn 1974). Lexicographic choice ranks the attributes of a house from most to least important: for example, it may be most important that a house is within budget, then that it has the requisite number of bedrooms, and so on. Lexicographic choice then takes the option which ranks best on the most important attribute. If there is a tie, lexicographic choice compares the remaining options on the second attribute, proceeding in this way until one option remains.

A common objection to lexicographic choice is that it allows small differences along a single dimension to decisively influence decisions. If one in-budget house has four bedrooms but there are five more in-budget houses with three bedrooms, then lexicographic choice eliminates these five remaining houses without considering their other attributes which may make up for the missing bedroom.

A common bugfix is to decide by *semi-lexicographic choice*, in which small differences between options along any given attribute are ignored. But semi-lexicographic choice has a flaw that lexicographic choice does not: it sometimes produces intransitive preferences or decisions across decision problems (Tversky 1969). Adherents of the Standard Picture would hold that on these grounds, lexicographic choice should be preferred to its semi-lexicographic cousin. Hence the Standard Picture is unable to explain how semi-lexicographic choice could constitute an improvement upon lexicographic choice in some choice situations. By contrast, most bounded rationality theorists hold that coherence is one among many features of a decision rule to be considered during strategy selection (Arkes et al. 2016). Decisionmakers facing close ties along a number of dimensions may make better decisions by choosing semi-lexicographically because this increase in decision quality can outweigh the vulnerability of semi-lexicographic choice to incoherence.

One of the central projects of this book is to develop an account of rational strategy selection (Lieder and Griffiths 2017; Marewski and Schooler 2011).[9] This account should say why it is often rational to inquire heuristically and shed light on how features of heuristics such as their compensatoriness and disposition to incoherence bear on their rationality.

2.5 Ecological rationality

Herbert Simon held that human decisionmaking is structured by a pair of constraints: our limited cognitive abilities, and the structure of the environment.

> Human rational behavior . . . is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor.
>
> (Simon 1990, p. 7)

In studying bounded rationality, it is easy to focus exclusively on limited cognitive abilities and neglect the structure of the environment. Most authors in the bounded tradition have followed Simon in thinking that this would be a grave mistake. We find ourselves located in certain environments, which pose special types of cognitive challenges and which structure the costs and benefits of possible responses. Rational cognizers use strategies designed to succeed in environments

---

[9] Actually that's not quite right. Strategy selection is a metacognitive process, and I will mostly be concerned with the rationality of inquiry strategies that result from strategy selection. My question is how factors such as compensatoriness and cue utilization bear on the rationality of strategies themselves, not the rationality of metacognitive processes that select strategies. See Chapter 8 for discussion.

like their own, not in all hypothetical environments. As a result, *rationality is ecological*, or environment-relative.

To illustrate the point, consider the defensive behaviors of adult robins (Lack 1943). As non-migratory birds, robins have a strong need to protect their territory against other members of the same species. One of the strongest triggers for aggressive behaviors by adult robins is the sight of a red breast feather. In their natural environment, this is an excellent way for robins to identify potential competitors: the primary distinguishing feature of the adult male robin is its prominent red breast. This strategy has other advantages as well: because juvenile robins lack red breast feathers, they are less likely to be mistakenly attacked. Nevertheless, we can manipulate the robin's environment to make their behavior maladaptive. For example, male robins have been known to vigorously attack a clump of red feathers but to avoid a stuffed replica of a male robin lacking red breast feathers.

Are the robin's defensive behaviors rational? In their natural environment, these behaviors are highly rational, combining a high rate of false positives with a low rate of false negatives. But in an environment populated by meddling scientists, red-headed humans or rivals who had learned to cover their breast feathers in mud, the robin's behaviors would no longer be rational. As a result, the right question to ask about the rationality of the robin's behaviors is not whether it is rational in any possible environment, but whether it is rational in the robin's environment.

But to say that rationality is environment-relative is not yet to say anything very precise. In what way does the environment bear on rational cognition? The bounded tradition is clear in taking ecological rationality to involve some type of match between the structure of the environment and the structure of an agent's cognition:

> Ecological rationality appears when the structure of boundedly rational decision mechanisms matches the structure of information in the environment.
>
>   (Todd and Gigerenzer 2012b, p. 3)
>
> In general, a heuristic is ecologically rational to the degree that it is adapted to the structure of the environment.   (Gigerenzer 2019)

This way of characterizing ecological rationality in terms of a match between cognition and the environment raises at least three questions that a good normative theory should answer.

First, what type of match is at issue in ecological rationality? A natural thought is that cognition matches its environment to the degree to which the agent's cognitive model used during deliberation reflects the actual structure of the environment. But this is manifestly not what the bounded tradition means by ecological rationality, since the bounded tradition holds that ecologically rational cognizers often

ignore most information available to them, as in the case of the robin. Perhaps the match between cognition and the environment is a matter of reasoning in ways that are *fitting* given the structure of the environment. But there is no evidence that key texts in the bounded tradition are concerned with fittingness.

I think that many authors in the bounded tradition are concerned with the question of how agents *ought* to cognize. On this interpretation, to say that an agent's cognition is ecologically rational is to say that she cognizes as she ought, given the structure of her environment. For example, here is how the editors of a recent anthology on ecological rationality characterize the question of ecological rationality.

> *Ecological rationality's question:* Given a problem in an environment, which strategies *should* humans rely on when optimization is not feasible?
>
> (Todd and Gigerenzer 2012b, p. 489, emphasis in text)

This interpretation may seem to involve a shift away from questions about rationality to questions about how agents ought to cognize. To meet this worry, I will argue that rationality is deontic: agents are rationally required to ϕ just in case they ought to ϕ. Although this interpretation leaves room for a conceptual distinction between rational and deontic questions, it explains why bounded rationality theorists are comfortable shifting between the coextensional questions of how agents are rationally required to cognize, and how they ought to cognize.

My deontic interpretation of ecological rationality raises a second question: what is the right deontic theory to capture mainstream claims about bounded rationality? Thinkers in the bounded tradition often assume a pragmatic theory on which agents ought to act in the ways that will satisfy their goals. For example, Peter Todd and co-authors hold that:

> The success of simple heuristics is defined with respect to pragmatic goals in a particular environmental context.   (Todd et al. 2000, p. 378)

A similar pragmatic approach to bounded rationality was introduced to philosophers through the work of Stephen Stich (1990).

Some other remarks in the bounded tradition suggest an adaptationist normative theory. We have already seen that Gigerenzer holds ecological rationality to be a matter of the degree to which a cognitive strategy is adapted to its environment, and this adaptationist language is reflected in other key statements of ecological rationality:

> Human reasoning and behavior are ecologically rational when they are adapted to the environment in which humans act.   (Rieskamp and Reimer 2007, p. 274)

While I concede that there is some evidence for an adaptationist reading of the bounded tradition, I think it is uncharitable to read the tradition as attributing any deep normative force to evolutionary adaptation given the availability of other, more normatively defensible readings. However, I will try to push the tradition away from pragmatism and toward consequentialism. This reflects my view that rational agents should aim not to achieve the aims they believe to be most important, but those aims which are actually most important.

A third question about ecological rationality concerns the objects to which it applies. Some statements of ecological rationality suggest that it applies to environment- and problem-tokens. For example, we have seen that Gigerenzer and Todd hold that ecological rationality applies at the level of a given problem and a given environment. But in practice, the bounded tradition usually asks questions about the performance of abstractly specified types or mechanisms of reasoning across a range of problems constituting a single environment. We should not be too quick to dismiss this focus as an artifact of research methodology. Consider, for example, the recent history of process reliabilism. Many process reliabilists hold that token processes of inquiry should be evaluated by their relationship to the types of inquiry processes that they instantiate, and to the reliability of those process types across an environment of similar problems. Process reliabilists make this move in order to divorce the success or failure of processes on a given occasion, which may be due to luck, from the success of those processes across a range of similar problems, which is both less luck-prone and perhaps also a more appropriate level at which to model human learning and decision processes.

If this is right, then a good normative theory should tell us whether ecological rationality applies to process- and environment types, tokens, or both. I will leave this question open by developing an account that has the flexibility to answer both types of questions, but at the same time, I will hold that questions about process- and environment types are often more helpful and stay closer to the normative questions that our theories are designed to answer.

Before concluding, we should note that ecological rationality, like many other aspects of the bounded tradition, is as much a methodological thesis as a normative claim. To say that rationality is ecological is to say not only that the structure of the environment matters to rational cognition, but also that models of rational cognition should strive as far as possible to incorporate the structure of the environment. We cannot ask of a heuristic whether it is rational or irrational simpliciter since most heuristics perform well in some environments and poorly in others. We should always ask instead: in what sorts of environments does a given heuristic strategy perform well or badly?

This last methodological thesis is one of the deepest disagreements between the bounded tradition and the heuristics and biases program, which holds that cognitive environments are typically too complicated to tractably model. The

introduction to a leading anthology begins by conceding one and only one point to its opponents, and it is this one:

> There is, however, one [point] . . . to which researchers in the heuristics and biases tradition must plead "no contest" or even "guilty." This is the criticism that studies in this tradition have paid scant attention to assessing the ecological validity of heuristic processes . . . This Herculean task has not attracted researchers in the heuristics and biases tradition; the focus has been on identifying the cues that people use, not on evaluating the overall value of those cues.
>
> <div align="right">(Gilovich and Griffin 2002, p. 3)</div>

Gilovich and Griffin have a point: modeling environmental structure is hard. But researchers in the bounded tradition protest that neglecting the structure of the environment is a mistake nonetheless, because it makes ecologically rational behaviors appear irrational by neglecting to specify the environmental conditions which rationalize them. Once we know how a heuristic works, we can design artificial environments in which the heuristic breaks down, but that fact alone tells us nothing about the ecological rationality of the heuristic itself. This is the crux of the bounded tradition's complaint against the heuristics and biases program:

> If we study biases without analyzing the structure of their environment, we can end up proposing processes that generate cognitive fallacies where none actually exist.   (Todd and Gigerenzer 2012b, p. 81)

This brings us to the last claim made by the bounded tradition: when viewed in the proper light, many apparent irrationalities are nothing of the sort.

2.6 Vindicatory epistemology

If we are open to moving beyond the Standard Picture of rationality, then the normative status of apparent cognitive biases becomes an open question. As we saw in Chapter 1, many biases are simply defined to be deviations from the Standard Picture, and many others rely on norms and intuitions closely allied to the Standard Picture. If we reject the Standard Picture as a normative theory, then the fact that patterns of cognition deviate from Standard Picture norms does not settle their rational status.

In this chapter, we have seen where normative theory should look to assess the rationality of apparent biases. Bounded rationality is process-focused, so we should focus not on the rational status of the outcomes of cognition such as belief, intention, and preference, but instead on the rational status of the processes

that produced them. Because cognition is often heuristic, a good place to start is by reassessing the rationality of heuristic cognition. And because rationality is ecological, we should ask of each heuristic not whether it is rational simpliciter, but rather whether it is used in an ecologically rational manner.

The program of *vindicatory epistemology* aims to recast apparent instances of biased cognition as the result of rational inquiry or metacognition. Vindicatory epistemology is sometimes parodied as the Panglossian view that all human cognition is as rational as it can be. For example, in 1981 Kahneman famously complained that his opponents see only two types of errors: "pardonable errors by subjects and unpardonable ones by psychologists" (Kahneman 1981, p. 340). This complaint was aptly directed at early vindicatory theorists, who often raised unfair complaints about experimental methodology and proposed ad hoc rationalizing explanations unsupported by new data or analyses. But in modern times, the vindicatory program has become increasingly able to deliver principled, empirically plausible rationality explanations (De Neys et al. 2013; Lieder et al. 2018; Oaksford and Chater 1994). As the vindicatory program has progressed, attitudes toward the program have softened. As an example, I return to Gilovich and Griffin:

> Researchers in [the heuristics and biases] tradition clearly share a set of assumptions: the ecological validities are probably high, the heuristics are generally useful, but common and profoundly important exceptions are to be found. (Note how this summary could be applied to the "fast and frugal" decision heuristics discussed . . . by Gigerenzer et al, despite the apparent opposition between the ecological rationality movement and the heuristics and biases perspective.)
>
> (Gilovich and Griffin 2002, pp. 8–9)

In this way, vindicatory epistemology cements the convergence between the heuristics and biases program and the bounded rationality camp by showing that many prominent heuristics are ecologically rational.

The last step needed for vindicatory epistemology is a principled normative theory by which the ecological rationality of heuristic cognition can be judged. This will avoid the charge that vindicatory theorists are choosing ad hoc and implausible normative standards after reviewing the data in order to make human cognition appear more rational. In Chapter 7, I will show how my preferred normative theory underwrites two well-known vindicatory projects. And importantly, the theory underwrites these and other vindicatory projects without change. Like many theorists in the bounded tradition, I think not only that human cognition is often ecologically rational, but also that existing vindicatory attempts have gone a good way toward demonstrating the ecological rationality of human cognition. I want to develop a theory that explains why these attempts are successful.

2.7 The way forward

In this chapter, we have met five key claims made by the bounded tradition: that bounds matter; rational assessment should be process-focused; heuristic cognition is often rational; rationality is ecological; and many apparent cognitive biases can be vindicated as instances of rational cognition. We have also seen the need for a normative theory to unpack the contents of these claims and to assess their correctness and implications.

In particular, we have seen that a good normative theory should do most or all of the following. It should clarify the relevance constraint by saying how bounds impact rational cognition, and meet the methodological constraint of incorporating resource bounds into all levels of normative theorizing. It should allow processes and their outcomes to be separable objects of normative assessment, while at the same time giving grounds for a focus on processes in normative theorizing. It should give an account of strategy selection which shows why heuristic cognition is often rational. It should say how the structure of the environment makes a difference to rational cognition. It should take a deontic interpretation of rationality. And it should capture the success of recent work in vindicatory epistemology without redescription or change.

In Parts 2–3, I develop and defend a normative theory designed to do just that. Part 2 outlines the theory, and Part 3 gives three arguments for its correctness. In Part 4, I show how my theory fits the job description outlined in this chapter. I argue that it performs most of the tasks set out in the previous paragraph and leaves room to make headway on the rest. I also show how this theory accommodates the Standard Picture as an approximately correct view about substantive rationality while often rejecting the Standard Picture as a theory of procedural rationality.

PART II
NORMS OF INQUIRY

3

An account of rational inquiry

3.1 Introduction

Part 1 introduced bounded rationality and the Standard Picture, using five key normative claims to characterize the bounded rationality approach. We saw that a normative theory of rational inquiry is needed to unpack the contents of these claims as well as to assess their correctness and implications. In Part 2, I construct a theory of rational inquiry. That theory will be defended and applied in Parts 3–4, respectively.

This chapter introduces the three main elements of the reason-responsive consequentialist view (RRCV) that I defend: a global consequentialist account of norms governing features of agency (Section 3.2); a reason-responsiveness conception of rationality (Section 3.3); and an information-sensitive account of deontic modals (Section 3.4). I also introduce some smaller commitments that will be integral to the application and defense of the RRCV (Section 3.5). An objection that arises in this discussion is that a consequentialist approach to rational inquiry is objectionably non-epistemic. I address this objection in Chapter 4.

3.2 Consequentialism: direct and global

3.2.1 Direct consequentialism

Early consequentialists were act consequentialists. They held that an action is right just in case it is best. But act consequentialism is an incomplete normative theory insofar as we are interested not only in the rightness of actions, but also in the rightness of other evaluands such as rules, motives, character traits, and legal systems. How should consequentialists assess these evaluands?

An initially popular proposal was *indirect consequentialism*. Indirect consequentialists apply consequentialism directly to evaluate the rightness of some privileged classes of evaluands. For example, motive consequentialists (Adams 1976) hold that the right motives to have are the best motives to have, and rule consequentialists (Brandt 1959; Harsanyi 1977) hold that the right decision rules to use are the best decision rules to use. Indirect consequentialists then apply consequentialism indirectly to determine the rightness of other evaluands in terms of their relationship to evaluands in the privileged class. Simple forms of rule- or

motive-consequentialism hold that the right actions are those that would result from the best set of rules or motives. In this way, indirect consequentialism marries the deontic status of some evaluands, such as actions, to the status of other evaluands, such as motives or decision rules.

The popularity of indirect consequentialism began to wane as cases emerged in which the deontic status of actions comes apart from the deontic status of the rules, motives and other privileged evaluands which produced them. For example, consider a baker who deals honestly with her customers because she is afraid of being caught. We would like to say that this baker acts rightly in dealing honestly with customers, but has the wrong set of motives. Now consider a baker who loves her family so much that she saves ten loaves of bread every day for her family. Suppose that the baker and her family would be better off if she sold nine of these loaves, but that the baker could only sell these loaves if she loved her family less. Here we would like to say that the baker acts wrongly in saving the loaves, but has the right set of motives. In this way, the rightness of motives and actions come apart. Wrong motives can produce right actions, and right motives can produce wrong actions. Similar examples can be generated to put pressure on other indirect approaches.

At this point, it will help to extend some terminology due to Shelly Kagan (2000). There are various *evaluative focal points* such as actions, motives, and decision rules at which we can apply normative terms. *Level tension* arises when the same normative term applies with different valences across causally related focal points (Thorstad 2021). The baker who saves too much bread for her family has the right set of motives but does the wrong thing as a causal result of these motives. And the baker who deals honestly with customers out of fear has the wrong motives but does the right thing as a causal result of these motives. *Indirect* normative theories tie the normative status of some focal points to the normative status of others in a bid to reduce level tension. By contrast, *direct* normative theories apply standards of correctness directly at each focal point.

An advantage of direct normative assessment is that it captures our complete normative reaction toward agents such as the baker. Julia Driver has argued that we feel an evaluative tension toward agents caught in level tension (Driver 2001). Our stance toward them is neither unequivocally positive nor unequivocally negative, but rather torn and conflicted. We feel at once as though we should praise the baker for her love and curse her for her stubbornness in saving so much bread. Direct normative assessment captures this evaluative tension in the natural way. We feel conflicted toward the baker because she has done something right, namely loving her family, and something wrong, saving too much bread.

A direct approach to normative theorizing is compatible with a number of deontic theories, as well as with a mixture of deontic standards across levels. For example, you might think that agents ought to believe a proposition just in case it

is supported by their current evidence, but that agents ought to inquire about what to believe using whichever strategies are best. This approach mixes an evidentialist account of right belief with a consequentialist theory of right inquiry.

However, direct consequentialism insists that one and the same deontic theory applies at each focal point. The right motives are the best motives and so too, the right action is the best action. This view delivers the correct verdicts in both versions of the baker case. The baker who deals honestly with customers from fear of discovery acts optimifically but from non-optimific motives. She would be motivated to steal even if, as is typically the case, it would be better not to steal. And the loving baker who saves too much bread because she loves her family acts non-optimifically from optimific motives. It would be better to sell the bread, but a disposition toward moderate monetary losses is a small price to pay for the benefits of a loving motive set.

Direct consequentialism is the starting point for a broader *global consequentialist* program. By developing direct consequentialism into a mature brand of global consequentialism, we will be in a better position to understand the nature and motivations for a consequentialist theory of rational inquiry.

3.2.2 Global consequentialism

Global consequentialism takes its inspiration from Derek Parfit. Here is how Parfit describes consequentialism:

> There is one ultimate moral aim: that outcomes be as good as possible. *Consequentialism applies to everything.*   (Parfit 1984, p. 24, emphasis added)

What does it mean to say that consequentialism applies to everything? On a natural reading, Parfit and other global consequentialists make the direct consequentialist claim that the right and the good coincide at each focal point. That is how Philip Pettit and Michael Smith describe the global consequentialist position:

> Global consequentialism identifies the right x, for any x in the category of evaluands—be the evaluands acts, motives, rules, or whatever—as the best x, where the best x is that which maximises value.   (Pettit and Smith 2000, p. 121)

This yields a first-pass statement of global consequentialism:

> **Global consequentialism, first pass:** For all X:
>
> **(Deontic-Evaluative Bridge)** X is right if and only if X is best.
> **(Bestness as Maximization)** X is best if and only if X maximizes value.

In this section, I propose three modifications to this first-pass account aimed at developing a version of global consequentialism suitable to ground a consequentialist account of rational inquiry.

An immediate objection to this account is that it neglects a crucial consequentialist stipulation: that all value is to be promoted (Pettit 1989). The claim that all value is to be promoted contrasts with the nonconsequentialist view that values are sometimes to be honored rather than promoted (Sylvan 2020). For example, suppose that if I tell a lie today I can spare several of my employees from lying in my stead. A first-pass consequentialist view says that I should lie because my lying promotes more overall truth-telling, whereas a first-pass deontological view says that I should not lie because in lying I dishonor the truth.[1] Or to take an example from recent epistemology, many epistemologists hold that I should refuse the 'epistemic bribe' of forming a false belief as a causal means to forming a large number of true beliefs (Berker 2013; Greaves 2013). A natural explanation for this claim is that forming a false belief fails to honor truth, even if it is used as a causal means to promote true belief.

Perhaps the distinction between promoting and honoring value is what Pettit and Smith had in mind by Bestness as Maximization. It was, after all, Pettit himself who introduced the distinction between promoting and honoring value. Otherwise, it is hard to see what Bestness as Maximization comes to.[2] At this point, a good way to improve upon our first pass is to replace Bestness as Maximization with the relevant form of the claim that all value is to be promoted.

Global consequentialism, second pass: For all X:

(**Deontic-Evaluative Bridge**) X is right if and only if X is best.
(**Promotion**) The value of X is determined by the goodness promoted by X.

Here I read Promotion broadly to allow different understandings of the promoting relation, as for example in the dispute between causal (Gibbard and Harper 1978; Joyce 1999; Stalnaker 1981) and evidential (Bolker 1966; Jeffrey 1965) decision theorists.

---

[1] Like most examples used to illustrate the distinction between honoring and promoting value, this example has several defects. For one thing, most consequentialists think that the primary value at issue is not honesty or even truth, but rather a combination of downstream instrumental values. For another, some consequentialists deny that I should lie in many cases of this sort, citing downstream negative effects of lying. The difficulty of illustrating the distinction between promoting and honoring value with a clean and simple example illustrates some of the underlying conceptual divisions between consequentialist and non-consequentialist views.

[2] Perhaps Bestness as Maximization was meant to distinguish Pettit and Smith's approach from satisficing consequentialism (Slote 1984) and other non-maximizing deontic theories. But that is not an issue to broach in this section.

Our second pass account is an improvement on the first pass, but it is nevertheless too strong. The second pass account follows Parfit in holding that consequentialism applies at all, or at least nearly all focal points:

> Consequentialism covers, not just acts and outcomes, but also desires, dispositions, beliefs, emotions, the colour of our eyes, the climate, and everything else. More exactly, [consequentialism] covers anything that could make outcomes better or worse.   (Parfit 1984, p. 27)

Here Parfit holds that consequentialism is a thesis both about features of agency, such as desires, dispositions, beliefs, and emotions, and also about other focal points which are not features of agency, such as eye colors and climates. By contrast, I think that consequentialism is best restricted to making claims about features of agency.[3]

When we consider focal points which are not features of agency, both components of the second pass account begin to break down. First, consider the Deontic-Evaluative Bridge. To say that the right climate or eye color is the best climate or eye color is, I think, a simple category mistake. Rightness and other deontic notions apply only to options, which even on the most liberal construal cannot fail to be features of agency. It may well be better for me to have blue eyes, but we would not want to say that I ought to have blue eyes or that it is right for me to have blue eyes. It is not clear what such claims could mean.

Second, consider Promotion. To say that the values of non-agential evaluands are determined by the goodness they promote is to make a perfectly meaningful, but highly controversial evaluative claim. Is a beautiful painting bad if it is adopted as a symbol by a group of fascists, or a sunset bad if it invokes the memory of happier days gone by? The motivation for Promotion is that as agents acting to shape the world, we aim to change the world for the better. Features of agency are good when they carry out this aim, and bad when they do not. But sunsets do not aim to change the world for the better. Sunsets do not aim at anything at all: they are sunsets, not agents. For this reason, it seems inappropriate to value sunsets by considering the goodness that they promote. Sunsets, unlike agents, are not in the business of promoting goodness.

This discussion suggests that we should restrict global consequentialism to features of agency:

Global consequentialism, third pass: For all features of agency X:

    (Deontic-Evaluative Bridge) X is right if and only if X is best.

    (Promotion) The value of X is determined by the goodness promoted by X.

---

[3] For defense and clarification see (Driver 2012).

What precisely are features of agency? On some views, only mental states like intentions or tryings are features of agency, in which case consequentialism is a theory about what we ought to intend or try to do. On many views, actions are also features of agency. It is an active area of debate whether beliefs and other doxastic attitudes are features of agency. Consequentialism is neutral on the question of what counts as a feature of agency. But the discussion of the Deontic-Evaluative Bridge suggests one reason to be liberal in what counts as a feature of agency: it does not make sense to apply deontic terms at focal points which are not features of agency. If believing is not something we do as agents, then there can be no question about what we ought to believe. Likewise, if heuristic inquiry is not a feature of agency when it is done unconsciously, then most of our inquiries will be inapt for deontic assessment.

The third pass statement of global consequentialism is not false, but it is incomplete. Global consequentialism is not just a deontic theory. If it were, then consequentialism would have nothing to say about rationality, virtue, and other important normative terms. But global consequentialists aim to expand the scope of consequentialist assessment to shed light on many such normative terms. For example, Julia Driver defends a consequentialist account of virtue:

> (**Aretaic-Evaluative Bridge**) A virtue is a character trait that produces more good (in the actual world) than not *systematically*.   (Driver 2001, p. 82)

On a natural reading, Driver's account proposes extending global consequentialism with a new bridge principle linking *aretaic* assessments of virtuous character to evaluative claims about the value produced by these character traits.

More generally, *expansionist* global consequentialism aims to strengthen global consequentialism with additional bridge principles bringing non-evaluative and evaluative claims together in quasi-consequentialist fashion. I say that these bridge principles are quasi-consequentialist because they will bear only a family resemblance to the Deontic-Evaluative Bridge. We should not expect to recover novel bridge principles of the form:

> (**P-evaluative bridge**): X has property P if and only if X is best.

because the existence of such bridge principles would show on a consequentialist view that property P is coextensive with rightness. Expansionist global consequentialism aims to show how many or most important normative terms, while remaining meaningfully distinct from rightness, can be analyzed in evaluative terms with a strong consequentialist component. For example, in the next section, I defend a bridge principle linking rationality to deontic facts. This bridge principle will be a strengthening of the Deontic-Evaluative Bridge.

How far can the expansionist program be pushed? On my view, there are clear examples of important normative terms, such as fittingness terms, which resist consequentialist analysis. The claim that an action is blameworthy is not connected in any obvious way to the claim that it would be beneficial to blame the actor. An action can be blameworthy even if the actor is unlikely to change their behavior and very likely to shout at anyone who blames them. Consequentialists have historically struggled to develop an account of blame, and if I am right that is because blameworthiness is not an especially consequentialist notion. Similar remarks apply to other fittingness categories. A person can be enviable even if it would not benefit anyone to envy them, for example, because envy is a negative and self-destructive emotion.

These barriers to expansionist global consequentialism should be unsurprising. We use normative terms to mark a rich variety of facts, some of which are closely tied to consequences and others of which have nothing much to do with consequences. But expansionist global consequentialists think that many of our most central normative terms such as 'ought' and perhaps also 'virtuous' can be brought within the scope of global consequentialism.[4] One of my projects in this book is to add rationality to the list of consequentialist normative terms. I carry out this project in the next two sections.

3.3 Reason-responsiveness

What is the relationship between rationality and reasons? To get a grip on this relationship, consider:

> **(Leo)** Leo is reading an informative philosophical book when he notices smoke coming from his lawn.

If he is rational, will Leo continue reading or investigate the smoke? Plausibly, Leo should investigate the smoke. Leo has more reason to investigate than smoke than to read philosophy, so that is what Leo is rationally required to do. This example suggests that agents are rationally required to do what they have most reason to do.

But doing what we have most reason to do is not sufficient for acting rationally. Consider:

---

[4] For my own part, I am a bit more reticent about consequentialist analyses of virtue. But this is not the place to air such squabbles.

(**Cleo**) Cleo is reading an informative philosophical book when she notices smoke coming from her lawn. Having learned from her crystal ball that smoking lawns mean leprechauns, Cleo rushes outside to investigate the smoking lawn.

Does Cleo act rationally in investigating the smoking lawn? Plausibly, she does not. Although Cleo has most reason to investigate the smoking lawn and that is what she does, Cleo does not investigate the smoking lawn in response to these reasons. Cleo investigates the lawn because she hopes to capture a leprechaun. This example suggests that acting rationally involves not only doing what we have reason to do, but also acting in response to the reasons which make our actions rational. Cleo would have acted rationally in investigating the smoke in order to stop a fire, but not in order to capture a leprechaun.

Many philosophers have taken cases such as Leo and Cleo to motivate a *reason-responsiveness* conception of rationality (Kiesewetter 2017; Lord 2018). On this conception, rationality consists in responding correctly to normative reasons. A bit more carefully:

> (**Rationality-Responsiveness Bridge**) For all agents *S*, times *t* and features of agency *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*.

Here I have stated the reason-responsiveness conception as a bridge principle to foreshadow my consequentialist ambitions. In this section, I motivate a reason-responsiveness conception of rationality (Section 3.3.1), then develop a global consequentialist interpretation of reason-responsiveness (Section 3.3.2).

3.3.1   Motivations

The reason-responsiveness conception has been defended at length by a number of theorists. For this reason, I confine myself to offering summary motivations for the view. In addition to its naturalness, there are three good motivations for adopting a reason-responsiveness conception of rationality.

First, the reason-responsiveness conception answers *Kolodny's challenge*: why be rational (Kolodny 2005)? On a reason-responsiveness conception, what we are rationally required to do is determined in the same way as what we ought to do: by the balance of possessed normative reasons (Lord 2018). Then Kolodny's challenge can be answered simply: in doing what is rational, we do what we ought and have most reason to do. By contrast, most competitors to the reason-responsiveness conception allow that we can be rationally required to X, but ought or have most reason to perform some other X′ instead. Then Kolodny's challenge sharpens to the following: why do what is rational instead of what we ought or have most reason

to do? This question is very difficult to answer in a way that preserves the authority of rationality.[5]

Second, the reason-responsiveness conception explains why normative theorists often slide between talk of what is rationally required and talk of what we ought to do. To illustrate the point, I made this slide myself in the discussion of Leo and Cleo, and I will wager that few eyebrows were raised in the process. While it is important to clearly separate oughts from rational requirements in our most careful moments, often we can pass freely between the two notions. Why is this so?

On a reason-responsiveness conception, the slide between oughts and rational requirements is not a deep mistake, because oughts and rational requirements are coextensive. That is,

> **(Rationality is Deontic)** For all agents S, times t and features of agency X, S is rationally required at t to X if and only if S ought at t to X.

Roughly the claim that Rationality is Deontic holds because what you are rationally required to do and what you ought to do are determined in the same way: by the balance of normative reasons.[6] Because Rationality is Deontic, when we are only concerned with the extensions of oughts and rational requirements, we can use either term without fear of error.

The reason-responsiveness conception also explains why, despite the frequent permissibility of the slide between ought- and rationality-talk, ought and rationality remain importantly distinct normative categories. On a reason-responsiveness conception, rationality is a strictly more demanding status than doing what we ought. Reason-responsiveness requires agents to do what they ought in response to the reasons for which they ought to do it. In this way, rationality as reason-responsiveness tracks a central type of normative status lacked by agents such as Cleo, who do what they ought but do so for the wrong reasons.

Third, the reason-responsiveness conception explains when and why irrationality defeasibly warrants blame. Many agents who act irrationally have done something they should not do, and agents are defeasibly blameworthy for doing what they should not. At the same time, some agents act irrationally but do what

---

[5] Some theorists bite the bullet here and concede that rationality is not authoritative (Broome 2013), or that epistemology studies non-authoritative reasons (Maguire and Woods 2020). I hope to show that there is a better alternative.

[6] The claim that Rationality is Deontic has the following troubling features: almost everyone agrees that it follows from a reason-responsiveness conception, but not everyone agrees on why it follows, and every convincing derivation is quite long (Lord 2017). Here is a brief sketch of my favorite derivation, but readers are welcome to substitute their own. On a reason-responsiveness conception, S is rationally permitted at t to X iff X-ing is a correct response to the reasons S possesses at t. X-ing is a correct response to the reasons S possesses at t iff S has sufficient reason at t to X, so pushing through we get that S is rationally permitted at t to X iff S has sufficient reason at t to X. Hence S is rationally required at t to X iff S has decisive reason at t to X. Since agents ought to do all and only what they have decisive reason to do, the claim follows.

they should. These agents are not obviously blameworthy. For example, suppose that Cleo believes that she ought to investigate her smoking lawn. Although this belief is irrational, it is not obviously blameworthy, and the reason-responsiveness conception explains why. Blame is defeasibly warranted by doing what we should not, but Cleo holds the belief that she should hold.

This way of cashing out the connection between rationality and blame solves a difficulty in the theory of epistemic blame. Many epistemologists have held that there is no clear species of blame connected to irrational belief, and they have been motivated in part by agents such as Cleo whose beliefs seem irrational but not blameworthy (Doughtery 2012). But on a reason-responsiveness conception, this is a familiar insight that does not cast doubt on the parallel between blameworthiness of belief and action. We should never have expected to recover a sense in which irrational belief defeasibly warrants blame, because it is not irrationality but rather doing what we should not which defeasibly warrants blame. On this view, it could well be the case that beliefs are blameworthy under the same conditions and for the same reasons as actions and other features of agency. For example, we could hold that beliefs are blameworthy if they ought not to be held and are unexcused.[7]

Summing up, there are four motivations for a reason-responsiveness conception of rationality. The reason-responsiveness conception is a natural way to account for cases such as Leo and Cleo; answers Kolodny's challenge; explains the slide between rationality and ought talk; and explains connections between rationality and blame. Let us turn now to developing the reason-responsiveness view.

3.3.2 Developing the view

On a reason-responsiveness conception, rationality involves responding correctly to possessed normative reasons. Several notions in this statement require clarification.

First, what does it mean to *respond* to reasons? A popular view is that features of agency X respond to reasons R when R bears a specified explanatory relationship to X. For example, we might require that R non-deviantly causally produce (Arpaly and Schroeder 2015; McCain 2012) or sustain (Audi 1983) X. Other authors require actors to have specific representational mental states with contents involving R, for example, the first-order belief that R is true or the higher-order belief that R is a reason for X (Audi 1986). We could also require that agents take (Boghossian 2014) R as a reason to X in deliberation, or that in acting they are

---

[7] Of course, this is not to say that we cannot develop new conceptions of epistemic blame (Boult 2021; Brown 2020). The point is only that we are not forced to change our theories of blame to accommodate blameless irrationality. However, it is an advantage of my view that to account for these simple phenomena, we do not need to posit a novel kind of rationality bearing tightened connections to blame, nor to posit a new kind of epistemic blame.

aware of the fact that R is a reason to X (Korsgaard 2005). Some authors require agents to manifest knowledge or competence to use R as a reason to X (Lord 2018). And some authors impose several of these conditions at once (Audi 1986).

I don't want to settle debates about the nature of the responding relation here. What I have to say will be compatible with many leading views. For present purposes, I will use Z to denote the relationship which obtains between normative reasons R and features X of S's agency when in X-ing, S responds to R.

Second, what does it mean to respond *correctly* to reasons? This question can be partly answered without invoking consequentialism. To respond correctly to reasons is to do what we have sufficient reason to do, and in doing so, to bear Z to the reasons that we possess, whatever relation Z turns out to be.

> **(Responsiveness-Reason Bridge):** For all agents S, times t and features of agency X, S responds correctly to the normative reasons that she possesses at t in X-ing if and only if:
>
> (RB1) S has sufficient possessed reason at t to X, and
> (RB2) S's X-ing at t bears Z to the normative reasons she possesses at t.

The Responsiveness-Reason Bridge should be acceptable to theorists of most stripes, and (RB1) can be weakened to make room for personal prerogatives (Scheffler 1994), satisfying criteria of correctness (Slote 1984) and other natural ways of modifying or denying the Responsiveness-Reason Bridge.

To fully flesh out the Responsiveness-Reason Bridge, we need to fill in (RB1) with an account of what agents have sufficient reason to do. It is at this point that global consequentialism enters the picture. For the global consequentialist, all reasons are reasons to promote value, which implies:[8]

> **(Reason-Evaluative Bridge):** For all agents S, times t and features of agency X, S has sufficient possessed reason at t to X just in case S's X-ing at t is best.

Putting together the Rationality-Responsiveness Bridge, Responsiveness-Reason Bridge, and Reason-Evaluative Bridge gives a global consequentialist account of rationality:

---

[8] This inference will not go through on views allowing incomparability of reasons, or views which allow that we can have sufficient reason to do something even though we have more reason to do something else. Addressing these views goes beyond the scope of our discussion, although it may be possible to make suitable modifications to the account. Another worry might be that the value of X may be determined not only by reasons that S possesses but also by reasons that S does not possess. This worry will be addressed by the information-sensitive account adopted in the next section, on which value is assessed from the same informationally constrained perspective that determines the reasons that agents possess.

(Rationality-Evaluative Bridge): For all agents S, times t and features of agency X, S's X-ing at t is rational if and only if:

(RB1′) S's X-ing at t is best, and

(RB2) S's X-ing at t bears Z to the normative reasons she possesses at t.

The Rationality-Evaluative Bridge has the same form as the Deontic-Evaluative Bridge, with the addition of a reason-responsiveness condition (RB2). Hence an equivalent way of stating the Rationality-Evaluative Bridge is:

(Rationality-Rightness Bridge): For all agents S, times t and features of agency X, S's X-ing at t is rational if and only if:

(RB1″) S's X-ing at t is right, and

(RB2) S's X-ing at t bears Z to the normative reasons she possesses at t.

In this sense, the Rationality-Evaluative Bridge has a good claim to being classed as a global consequentialist account, because it gives an account of rationality which is a more demanding form of the consequentialist account of rightness.

A third question about the reason-responsiveness view is what it means to possess normative reasons. It turns out that there are sharp constraints on the accounts of reason possession that can be used to develop a reason-responsiveness account of bounded rationality. In the next section, I defend a view of reason possession that meets these constraints.

3.4 Possessed reasons: objective, subjective, and information-sensitive

3.4.1 Moving beyond objectivism and subjectivism

The Deontic-Evaluative Bridge says that a feature of agency is right just in case it is best. But from what perspective is the goodness of a feature of agency determined? This question has traditionally been answered in two ways. *Objective consequentialists* think that a feature of agency is best if it would actually have the best consequences. *Subjective consequentialists* think that a feature of agency is best if the agent in question believes it to have the best consequences.

By the Reason-Evaluative Bridge, the goodness of a feature of agency is determined by the reasons possessed by the agent, hence subjective and objective consequentialism can be reformulated as claims about the nature of possessed reasons. In these terms, objective consequentialists hold that:

(Objectivism about Possessed Reasons): *S* possesses *R* at *t* as a reason to *X* if and only if at *t*:

  (OP1) *R* is true, and
  (OP2) *R* counts in favor of *X*.

Subjectivists hold, by contrast:

(Subjectivism about Possessed Reasons): *S* possesses *R* at *t* as a reason to *X* if and only if at *t*:

  (SP1) *S* believes that *R* is true, and
  (SP2) *S* believes that *R* counts in favor of *X*.

These statements of Objectivism and Subjectivism about Possessed Reasons are compatible with many views about the metaphysics of reasons, for example with views on which reasons are propositions (Fantl and McGrath 2009) or facts (Littlejohn 2012). Other views will give rise to natural reformulations: for example, if reasons are mental states (Davidson 1980) then it will be the content of *R*, rather than *R* itself which figures in each statement.

The dominance of objectivism and subjectivism presents a problem because both views face difficulties as accounts of bounded rationality. The problem with objectivism is that it ignores one of the most central bounds in the cognitive lives of most agents, namely informational bounds. Just as we are computationally bounded agents, limited in our ability to perform mental computations, so too we are informationally bounded agents, limited in our access to information bearing on those computations. Theories of bounded rationality can no more ignore informational bounds than they can ignore computational bounds. To do otherwise risks contradicting the claim from Chapter 2 that bounds matter to rational cognition.

Traditional arguments against objectivism have stressed exactly this point: objectivism does not capture the normative importance of informational bounds. Consider a case due to Frank Jackson.

(Jill) Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug.

(Jackson 1991, pp. 462–3)

Objective consequentialism says that Jill should give the patient whichever of drugs B or C will completely cure the patient, since that is the action that would have the best consequences. And that is a tough pill to swallow. For the objective consequentialist, Jill ought to act as though she had full information about the consequences of candidate actions. When treating patients, she ought to give them the treatment which would in fact be best, even if she has every reason to expect that another treatment will have better results. When playing poker, she ought to hold on to a terrible hand if, in fact, the cards yet to be flipped would be favorable. And that does not seem right. Jill does not have complete information and theories of bounded rationality should take this fact into account.[9]

It might seem that subjectivism correctly reflects the normative impact of informational bounds since those bounds are incorporated into agents' beliefs. But the beliefs of bounded agents do not always incorporate relevant information very well, even if those agents are rational. One problem is that the beliefs of boundedly rational agents may ignore relevant information. We saw in Chapter 2 that in low-stakes contexts, rational agents often form beliefs using a small subset of available information in order to best balance accuracy and effort in cognition. When those beliefs are not supported by the totality of available information, subjectivism says that agents have reason to do what their information suggests they should not.

Another problem is that boundedly rational agents may respond to some reasons R while failing to have any beliefs at all either about the truth of R, or about the status of R as a reason. In such cases, subjectivism suggests that bounded agents do not possess R as a reason, even when they respond systematically and correctly to R.

To illustrate the second problem, suppose you are memorizing vocabulary words using flashcards. You would like to set aside the cards you have memorized well and keep studying the cards you have not memorized well. Suppose that C is a card that you have spent more time memorizing than other cards. Should you expect C to be better- or worse-learned than other cards in the deck? The answer is that C is probably less well-learned than other cards. There is an inverse relationship between study time and learning because more time is devoted to studying the most difficult cards (Koriat et al. 2006).

Let R be the fact that C has been studied for an abnormally long time, and let X be the action of devoting additional study to C. Most agents respond correctly to R in self-directed study. That is, we judge that C has been less well-learned than

---

[9] Objectivists have several replies available. For example, they can say that while Jill ought to do these things, she should not be blamed for failing to do them, and in fact she could be blamed for exposing herself and her patients to unnecessary risks. But objectivists have struggled to develop a systematic and plausible account of how informational limitations bear on what agents ought to do.

other cards and as a result perform X.[10] But do we believe that R is a reason to X? There is some evidence that we do not. For example, we often do not apply the inverse relationship between study time and learning in assessing how well others have learned (Koriat and Ackerman 2010). Many theorists have taken this finding to imply that we respond to R by responding to features of our experience such as felt fluency of recall, but that we respond directly to these cues without activating metacognitive beliefs, for example, the belief that R counts in favor of X (Koriat 1997). Then there is no good reason to suppose that we have the belief that R counts in favor of X. It is not always cognitively efficient to back up our metacognitive practices by forming beliefs about the goodness of those practices, at least when the beliefs are unnecessary to implement the practices.

If this is right, then subjectivism gives the wrong result. Subjectivism says that R is not a possessed reason to X because we do not believe that R counts in favor of X. As a result, subjectivism may well say that we should not do X. But we would like to say that bounded agents act rightly in X-ing, and they act rightly by responding to R on the basis of sensitivity to metacognitive cues in their experience of studying C which are appropriately related to R. R is a true proposition that counts in favor of X-ing, and our experience provides us with ample evidence that R is true and counts in favor of X-ing. The fact that R is not explicitly believed is no hindrance to the ability of bounded agents to respond correctly to R as a reason, nor is it an excuse for bounded agents to ignore R.

The upshot of this discussion is that we need a new account of reason possession. Neither objectivism nor subjectivism will do for theorizing about bounded rationality. How should bounded rationality theorists think about possessed reasons?

3.4.2 Information-sensitivity

Can we find a middle-ground between the objective and subjective ought which is appropriately sensitive to available information? Indeed we can. We can take the correctness of Jill's actions to be a function, not of her beliefs about consequences nor of the consequences themselves, but rather of information about consequences available to Jill. Jill ought to take the action with the best expected consequences, where the probabilities in question are derived from the available information.

This view belongs to a family of *information-sensitive semantics* for deontic modal vocabulary (Charlow 2013; Kolodny and MacFarlane 2010; Silk 2014). Like any family, we have our disagreements, and I will try to present information-sensitive semantics in a general way that the whole family can

---

[10] Actually, we may also lack the belief that C has been less well-learned, instead choosing to study C based on *noetic feelings* during metacognition (Arango-Muñoz and Michaelian 2014; Proust 2013). But I do not want to rest my case on this stronger and more controversial claim.

accept.[11] On my view, information affects an agent's obligations by contributing a probability function. That view should be agreeable to most epistemologists who accept the existence of evidential probabilities. Otherwise, readers are invited to insert their favorite story about objective probability. Perhaps the story is broadly frequentist, against a class of environments similar to our own in which all available information remains true. But some theorists take information to bear directly on what ought to be done, without the mediating influence of a probability function (Kolodny and MacFarlane 2010). Others replace probability functions with weaker objects such as imprecise probabilities.[12] Some theorists grant that information contributes a probability function, but adopt a different decision rule than expected utility maximization.[13] For generality, I present information-sensitive semantics without making any assumptions about the type of object contributed by bodies of information or about the decision rule relating that object to what agents ought to do. There is also some controversy about which body of information is relevant to assessing Jill's actions. I take a more committed stance on these issues elsewhere (Thorstad manuscript a).

Objective and subjective consequentialists share the assumption that deontic modals such as 'ought' are evaluated at worlds.[14] Many non-consequentialist views make the same assumption. The differences between these accounts of deontic modals can be captured in terms of *deontic selection functions d*, mapping a world w in which agents must act to the set $d(w)$ of ideal worlds capturing how the history of that world might continue if the agent acts appropriately. Most of these accounts accept a semantics on which claims ϕ about what agents ought to do are true at a world w just in case they are true throughout the deontically ideal worlds $d(w)$. Writing $\Box_d$ for the deontic modal 'ought,' read in accordance with selection function d, the common assumption is

$$w \models \Box_d \phi \text{ iff } (\forall w' \in d(w))(w' \models \phi) \tag{3.1}$$

Objective consequentialism results from taking $d(w)$ to be the worlds in which the agent's actions have the best consequences. Subjective consequentialism results from taking $d(w)$ to be the worlds in which the agent's actions maximize subjective

[11] Perfect family harmony is, of course, unobtainable. Some authors think that information-sensitivity can be captured by a Kratzer-style semantics, which is not the type of semantics that I present. But although this is a significant disagreement from the perspective of formal semantics, it does not reflect significant disagreement about the underlying normative relevance of information to deontic obligations.

[12] Although imprecise probabilities take up the lion's share of philosophical attention to generalizations of probability theory, there are many good options here. See Halpern (2003) for an overview.

[13] See Carr (2015) and Charlow (2016) for arguments that semantics should be flexible on the matter of decision theory, and two different approaches to spelling out that flexibility.

[14] This is a simplified presentation of the Kratzer-style semantic assumptions that have dominated semantic and normative discourse about deontic modality until recently. See Kratzer (1981).

expected value. Non-consequentialist theories may take $d(w)$ to be the worlds in which no deontic requirements are violated, or in which the agent acts virtuously.

An information-sensitive reading proposes that all of these theorists have made a mistake. Deontic modals are sensitive not only to how things actually are, but also to what information i is available. So deontic modals should be evaluated relative to a pair $<w, i>$ of a world and an information state. And the deontically ideal worlds are a function $d(i)$ of available information, rather than background facts.[15]

$$<w, i> \models \Box_d\phi \text{ iff } (\forall w' \in d(i))(<w', i> \models \phi) \qquad (3.2)$$

On the reading I propose, the subjective consequentialist deontic selection function was on the right track. The set $d(i)$ is indeed the set of worlds in which the agent takes some action maximizing expected utility. But the probabilities used to calculate expected utilities do not derive from an agent's own beliefs. They are instead relative to available information. For example, we might hold that the deontically ideal worlds are those in which the agent's actions maximize expected utility given her evidential probabilities. Then we recover a natural semantics on which agents ought to do whatever maximizes evidentially expected utility.

That is what the information-sensitive view says. Why should we accept it?[16]

3.4.3 Support for information-sensitivity

There are at least four good reasons to accept the information-sensitive view. First, unlike subjectivism and objectivism, information-sensitivity builds informational bounds directly into the traditional consequentialist story about how agents ought to act, for example by determining the evidential probabilities used to calculate expected value. By contrast, objectivism makes informational bounds irrelevant to how agents ought to act, and subjectivism makes informational bounds relevant only insofar as they are reflected in an agent's beliefs. When agents' beliefs imperfectly reflect available information, an information-sensitive semantics comes apart from subjectivism in letting the full importance of available information bear on how agents ought to act.

Second and relatedly, information-sensitivity gets the right verdict in problem cases.[17] Consider first Jill, who may relieve all her patient's symptoms with Pill A,

---

[15] Some theorists make the deontically ideal worlds $d(w, i)$ sensitive to worlds as well as bodies of information (Silk 2014). This is not my view.

[16] In addition to the advantages of the information-sensitive approach described in this section, the approach will later prove useful in clarifying and defending the claims that bounds matter (Section 8.2) and that rationality is ecological (Section 8.5). It will also illuminate one way in which my view of rational attitudes differs from the requirements placed on unbounded agents (Section 9.3).

[17] Information-sensitive semantics also famously shed light on other problem cases for objectivism and subjectivism, such as miner puzzles (Kolodny and MacFarlane 2010).

or else choose among Pills B and C, one of which will cure her patient and the other will kill them. Unlike objectivism, information-sensitive consequentialism correctly holds that Jill should administer Pill A because the evidentially-expected values of Pills B and C are dangerously low. And unlike subjectivism, information-sensitive views correctly continue to hold that Jill should administer Pill A if we imagine that Jill believes, against all evidence, that Pill B will bring on a leprechaun.

Now consider our flashcard memorizer, who has ample evidence that her oft-studied card C is likely to need further study. Given this evidence, information-sensitive views hold that the agent should study C. Unlike subjectivism, information-sensitive views do not forbid the agent from studying C if she has false beliefs or no beliefs at all about the relationship between learning and study. And unlike objectivism, information-sensitive views do not say that the agent should put down the card if, against all odds and unbeknownst to the agent, her last repetition etched the contents of C firmly into long-term memory. These seem like the right verdicts for a theory of bounded rationality to recover.

A third reason to opt for information-sensitive views is that they account for a range of novel sentences that traditional views struggle to explain. Consider, for example, *probabilistic-deontic conditionals* which mix probabilistic language in the antecedent with deontic language in the consequent (Cariani 2016). The probabilistic-deontic conditional

> If the sun is likely to explode tomorrow, you ought to feast tonight.   (3.3)

cannot be analyzed as a material conditional. On that reading, (3.3) comes out vacuously true, but so does:

> If the sun is likely to explode tomorrow, you ought to go to bed hungry.   (3.4)

An information-sensitive reading allows us to do better.

On this reading, probabilistic-deontic conditionals are *information-shifting* (Kolodny and MacFarlane 2010). The antecedent of the conditional adds a new proposition to the agent's information set: that the sun is likely to explode tomorrow.[18] To evaluate the conditional, we then test whether the consequent is true on the revised set of information. On this reading, (3.3) is true because feasting is an optimific response to the likelihood of death tomorrow. (3.4) is false, because fasting is not an optimific response to likely death.[19]

---

[18] More formally, Kolodny and MacFarlane (2010) hold that $<w, i> \models [\text{if } \phi]\psi$ iff for every maximal ϕ-subset $i' \subseteq i$ we have $<w, i'> \models \psi$ where a maximal ϕ-subset is i' such that $(\forall w \in i')(<w, i> \models \phi)$.

[19] Some readers with more ascetic temperaments may disagree with these verdicts. On an information-sensitive reading, these readers should donate their victuals to the rest of us.

Information-shifting semantics are increasingly popular readings of many conditionals, including probabilistic-deontic conditionals. But the success of this account relies on an information-sensitive account of deontic modals. If the deontic claim 'you ought to feast tonight' did not depend on available information, then its truth would be unaffected by updating on the information that the sun is likely to explode. So we cannot reap the explanatory benefits of information-shifting conditionals without taking their deontic consequents to be information-sensitive.

A fourth and final reason to go information-sensitive is that information-sensitive accounts of epistemic modal vocabulary have been enormously successful (Yalcin 2007). It is very natural to extend the same resources to cover deontic modals as well.[20] In fact, the semantic commonality between epistemic and deontic modals is already embedded in folk psychology through natural language. Although many modals such as 'ought' have both epistemic and deontic readings, these modal terms are not lexically ambiguous.[21] This means that ordinary language treats epistemic and deontic vocabulary as having a uniform underlying semantics. Epistemic and deontic readings of 'ought' are generated by different values of the same semantically relevant parameters. Hence if the epistemic ought is information-sensitive, then so is the deontic ought. Of course, the folk could be wrong in this regard. But the fact that we use the same lexical item to express both deontic and epistemic modal claims suggests that we take the same types of parameters to bear on the truth of each.

In this section, we have met four reasons to accept an information-sensitive account of deontic modals. Information-sensitivity builds informational bounds directly into deontic theory; accounts for problem cases; explains a range of semantic phenomena; and extends insights from the theory of epistemic modals. At this point, all three major components of the reason-responsive consequentialist view are in place: a global-consequentialist theory, expanded to cover at least rationality and rightness (Section 3.2); a reason-responsiveness account of rationality (Section 3.3), and an information-sensitive account of deontic modals (Section 3.4). Before defending and applying the account, it will help to introduce a few more commitments in order to strengthen and clarify the view.

[20] But see Yalcin (2016) for a cautionary note: weak necessity modals ('should,' 'ought') may lack genuine epistemic readings. This does not defeat the point: necessity modals ('must,' 'have to') and possibility modals ('may,' 'might') will still have dual readings.

[21] One way to see this is that modal vocabulary with dual epistemic and deontic readings is found in most major languages. Lexical ambiguity does not reproduce across language families.

3.5 Additional components of the account

In this section, I introduce three further components of the RRCV: a sharp *level separation* between questions about rational belief and inquiry; a *rich axiology* on which many intellectual and nonintellectual achievements bear final value; and a *consequentializing program* aimed at recovering detailed, plausible, and unifying explanations of normative phenomena.

3.5.1 Level separation

One of the major challenges facing a global consequentialist account of rational agency is that it threatens to generate significant pragmatic encroachment on belief. On a reason-responsiveness view, if beliefs count as features of agency then what we are rationally required to believe is what we ought to believe. And what we ought to believe is accounted for in consequentialist fashion:

> **(Rationality-Evaluative Bridge for Belief):** For all agents S, times t and propositions p, S's belief that p at t is rational if and only if:
>
> I S's belief that p at t is best, and
> II S's belief that p at t bears Z to the normative reasons she possesses at t.

The Rationality-Evaluative Bridge for Belief reduces questions about rational belief to questions about what is best to believe. And by:

> **(Promotion)** The value of X is determined by the goodness promoted by X.

what is best to believe is determined by the value promoted by our beliefs.

It is natural to think that our beliefs promote many types of value: not only intellectual values such as truth, knowledge, and understanding, but also non-intellectual values such as welfare and equality. If this is right, then the reason-responsive consequentialist view opens the door to the traditional pragmatist claims that it may sometimes be rational to have faith in God (Plantinga 1983) and humanity (Preston-Roedder 2013), or to think the best of our friends (Keller 2004; Stroud 2006) and ourselves (McKay and Dennett 2009; Rinard 2019b), even if these beliefs are not supported by the evidence. Here it is perhaps too mild to describe the threat as a type of pragmatic encroachment, in which traditionally non-epistemic factors impinge on a standard account of epistemic rationality. The threat is rather that we have adopted a thoroughgoing pragmatic, or rather consequentialist account.

This is a serious worry, and I will respond in detail to this worry in Chapter 4. For now, here is one way to soften the worry.

Direct consequentialism posits a sharp *level separation* between questions about rational belief and inquiry (Thorstad 2021).[22] It is one question what we ought to believe, and another question entirely how we ought to inquire about what to believe. Although consequentialists should accept a robust influence of practical considerations on questions about rational inquiry, they need not accept any influence of practical considerations on questions about rational belief.

Level separation is familiar from discussions of practical inquiry. An early objection to consequentialism was that consequentialist decisionmaking is too cognitively demanding: agents cannot always explicitly calculate expected values. The traditional consequentialist response is to sharply separate questions about practical inquiry and intention (Parfit 1984; Railton 1984). It is true that consequentialism requires agents to intend what is expectedly best. But it is not true that consequentialism requires agents to inquire by explicitly calculating expected utilities. If it would be better for agents to inquire in some other way, for example by using decisionmaking heuristics, then that is what they ought to do. It may, of course, turn out that rational heuristic inquiry sometimes produces intentions that are not expectedly best. When that occurs, we get a form of level separation between the rational status of intentions and inquiries: a rational process of inquiry can produce an irrational intention. But that fact is neither contradictory nor surprising, so it is no objection to the consequentialist view.

I think that consequentialists should say the same thing about the relationship between theoretical inquiry and belief as they say about the relationship between practical inquiry and intention. There is a robust influence of practical facts on questions about rational inquiry. You should, for example, often spend more time investigating important matters than unimportant matters. But this alone is no reason to posit any influence of practical considerations on questions about rational belief. It is, of course, true that the beliefs we form will be influenced by the inquiries that we undertake, and these inquiries are rationally responsive to practical considerations. But those practical considerations need not have any influence on the rational status of the resulting beliefs. If we are rationally required to make only the most cursory inquiries into some subject in order to conserve cognitive resources, then the result of rational inquiry may well be that we form an irrational belief on the subject in question. And the rationality of that belief may be determined entirely by evidential considerations.

---

[22] Level separation will feature heavily in my treatment of epistemic rationality in Chapter 4, as well as in my final account of process-focus (Section 8.3) and my preferred strategy for vindicatory epistemology (Section 7.4).

Now it might seem that a global consequentialist must accept the existence of practical reasons for belief. After all, beliefs can promote many types of value, and if beliefs are features of agency then reasons for belief are reasons to promote value.[23] However, consequentialists can make use of a number of traditional strategies for denying that the reasons in question are reasons for belief.[24]

Epistemologists have traditionally held that apparent practical reasons for belief cannot be genuine reasons for belief, for example, because they do not play the right role in motivating belief (Kelly 2002, 2003) or in deliberation about what to believe (Shah 2003, 2006). It is held that apparent practical reasons for belief are really reasons for some activity, such as getting ourselves to believe (Way 2012). For example, the importance of a comfortable retirement is not a reason to hold accurate beliefs about retirement, but rather a reason to get yourself to hold accurate beliefs about retirement through activities such as inquiry. This strategy rests on precisely the sharp level separation between questions about belief and inquiry that consequentialists suggest. Apparent practical reasons for belief are, on this strategy, genuine normative reasons for inquiry, but are not normative reasons for the beliefs that result from inquiry.

I think that this discussion provides ample motivation for adopting a sharp level separation between belief and inquiry, and I hope that many traditional epistemologists will be satisfied with the prospects for using level separation to avoid pragmatic encroachment on belief. Now of course, it may turn out that these traditional arguments fail, and that some apparent practical reasons for belief are genuine normative reasons for belief (Rinard 2019a; Steglich-Petersen 2011). If that is the case, then we will get pragmatic encroachment on belief. But here the culprit will not be global consequentialism, but rather the failure of traditional epistemological strategies for resisting pragmatic encroachment on belief. If those strategies fail, then all adequate views of rational belief will exhibit pragmatic encroachment.

---

[23] A road not taken in this discussion would be to deny the converse of Promotion: although all normative reasons are reasons to promote value, not all opportunities to promote value are normative reasons. I have not taken this road because I am not sure how to motivate this move in the context of rational belief while denying it in the context of rational inquiry, and also because there is a lot to be said for the roads most commonly traveled.

[24] Some care must be taken to translate this traditional discussion into terms compatible with the RRCV. For example, the claim that consequences are reasons for getting oneself to believe, rather than reasons for belief, could be phrased as the claim that it is inquiries rather than the belief states they produce which promote valuable consequences. Alternatively, traditional epistemologists might deny that there is any such thing as all-things-considered rational belief (Feldman 2000, 2002), or hold that they are talking about something else such as epistemic rationality (Firth 1956, 1959) or justified rather than rational belief (Goldman 1986; Lyons 2016; Siscoe 2021). Any of these moves would make traditional evidentialist theories compatible with the RRCV, although I do not want to pronounce on the advisability of any specific move.

3.5.2 Rich axiology

Many epistemologists think that intellectual achievements such as knowledge, understanding, and accurate belief bear significant value. Although some of that value is due to the practical role of our intellectual achievements in guiding action, a large part of the value of intellectual achievements is intrinsic to the achievements themselves. Humans are, after all, thinking creatures, and a full mind is every bit as much a part of a life well lived as a full belly.

Rudimentary forms of pragmatism will struggle to account for the intrinsic value of intellectual achievements. For example, if we take pragmatism to be the claim that the value of features of agency is determined by the non-cognitive benefits they promote, then there is no room left to accept that knowledge and understanding are also valuable in their own right. But consequentialists need not struggle at all to account for the intrinsic value of intellectual achievements.

One consequentialist strategy would be to begin with an impoverished axiology, such as welfarism, on which only welfare bears final value. We would then argue that the intrinsic value of intellectual achievements can be accommodated within a more impoverished axiology. For example, welfarists might say that just as social achievements such as friendship are an important constituent of well-being for social creatures such as ourselves, so too intellectual achievements such as knowledge are an important constituent of well-being for thinking creatures such as ourselves.

But consequentialism does not force us to make such reductive claims. Consequentialists can adopt any theory of final value that they wish. Many non-consequentialists adopt a *rich axiology* on which many things, such as welfare, knowledge, and friendship, bear final value. For my part, I am inclined to agree. Consequentialists can accept that knowledge and other intellectual achievements bear final value for exactly the same reasons that any other epistemologist would accept these claims. I will not argue for the adoption of a rich axiology, because I suspect that the impoverished route could say many of the same things as the richer approach about cases of interest. For example, when I claim that knowledge in its own right is a valuable state for humans to be in, the impoverished route will say that knowledge in its own right is an important component of well-being.

The next subsection will remind us why it is important for consequentialists to adopt a rich axiology, or at least to recover the intrinsic value of intellectual achievements from within a more impoverished axiology. Consequentialists have a consequentializing program. We aim to show how plausible claims about rational belief and inquiry can be recovered from within a consequentialist framework. The most challenging part of consequentializing is the recovery of traditional truth-directed epistemic norms. One traditional way to recover truth-directed norms is to claim on Cliffordian grounds that true belief is needed to promote valuable ends (Clifford 1877). We can stress that agents are generally in a better position

to promote valuable ends when they have correct views about how the world is and what effects their actions will have on the world. And we can stress that agents often fail spectacularly to promote valuable ends when they have incorrect views about the world and the effects that their actions will have on the world.

But Cliffordian arguments alone may not always be strong enough to recover truth-directed norms. Here it is helpful to supplement Cliffordian arguments with the claim that truth, knowledge and other intellectual achievements have value in their own right, in addition to their Cliffordian value in promoting pragmatic goals. We do not just seek knowledge in order to build ships, but also because knowledge is valuable. And that is just what a rich axiology allows us to say.[25]

3.5.3 Consequentializing

Like any theory, consequentialism has its share of counterintuitive consequences. On a first pass, consequentialism threatens to condone pushing fat men off of bridges, cutting up surgical patients to distribute their organs, and spurning friends in favor of strangers. Consequentializing, in its traditional sense, aims to show how consequentialism can deliver the intuitively correct verdicts in many of these cases.

In the Good Old Days, consequentializing was a simple affair. Consequentializers tackled purported counterexamples one at a time, telling detailed stories about how the verdicts in question could be recovered in an explanatory and non-arbitrary way using consequentialist tools. For example, traditional consequentializers argued that it may be wrong to cut up the one to save the five because doing so would undermine trust in medicine. And they argued that this story had explanatory advantages over its nonconsequentialist counterparts. For example, if it seems impermissible to cut up the one to save five, it seems permissible to cut up the one to save five million, and a traditional consequentializing story makes ample room for the sorts of tradeoffs needed to recover this verdict. An important feature of these traditional consequentializing approaches is that they did not try to recover every possible nonconsequentialist verdict. Faced with an interlocutor who insisted that it is impermissible to cut up the one to save five million, the traditional consequentializer would have told this interlocutor that she was mistaken. It is better to cut up one person in order to save five million, and so that is what we ought to do.

Recent consequentializing approaches have had broader aims. It is argued that all, or nearly all plausible normative theories can be replaced by a consequentialist

---

[25] More generally, a rich axiology will be useful in accounting for norms of clutter avoidance (Section 6.2) and friendship (Section 6.3), modeling rational anchoring and adjustment (Sections 5.6 and 7.5), and avoiding simple forms of pragmatism (Section 5.5).

equivalent that coincides with the original theory in all of its deontic verdicts (Dreier 2011; Portmore 2011; Seidel 2019). These newfangled consequentializing approaches are made possible by an expanded reading of traditional consequentialist mainstays like consequence and outcome, together with a liberalized theory of value. For example, a contemporary consequentializer might claim that it is wrong to cut up the one to save the five, because the outcome of cutting up the one would be a killing, and a killing is worse than five lettings-die. Faced with a deontologist who claims that it would still be wrong to cut up the one to save five million, this consequentializer need not tell the deontologist that she is wrong. The consequentializer could instead recover the deontologist's view by adopting an axiology on which any number of lettings-die is better than a single killing.

These newer consequentializing approaches have met with considerable resistance. I don't want to pronounce on the correctness of these challenges. But I do want to outline some of the skeptical challenges that have been raised for newer consequentializing approaches in order to show how my project avoids these challenges.

One objection that has been raised is that contemporary consequentializing programs are trivial (Betzler and Schroth 2019; Tenenbaum 2014). We might have hoped that consequentialism would tell us something specific about what it is right or rational to do. Certainly that is my aim in this book. But now the consequentializer promises to deliver nearly any verdict about rightness and rationality. If that is right, then consequentialism may not be much use in settling questions about rational inquiry.

A second objection is that contemporary consequentializers no longer put the *good* prior to the right (Schroeder 2007). This objection holds that when contemporary consequentializers talk about the good, they are no longer using the word in its ordinary sense. We can, of course, say of any difference between outcomes that it is a difference in the value of those outcomes. For example, we can say that an outcome is better for being a letting-die than for being a killing. But objectors have thought that it is not clear how claims like this could be true if goodness takes its ordinary meaning. It is perfectly intelligible that an act could be better for being a letting-die than for being a killing. But it is less clear why this makes the outcome of the first act better than the outcome of the second act if terms like goodness and outcome take their ordinary senses.

A third objection is the contemporary consequentializers no longer put the good *prior* to the right (Tenenbaum 2014). Many consequentialists think not only that we ought to do what is best, but also that we ought to perform this action because it is best. But as we liberalize the notion of doing what is best, it becomes increasingly plausible to cite other facts in explaining why an action ought to be performed. For example, suppose you think that an outcome is worse if it involves a killing than a letting-die. You might say that we ought to bring about the letting-die rather than the killing, because bringing about a letting-die is better than bringing about a

killing. That reading does put the good prior to the right. But you might also say that we ought to bring about the letting-die rather than the killing, because killing is wrong, or because we have a duty not to kill. And you might even add that a letting-die is better than a killing, because killing is wrong. That reading puts the right prior to the good. So it is not yet clear that contemporary consequentializing programs bear out the traditional consequentialist slogan that the good is prior to the right.

Fourth, it has been argued that contemporary consequentializers undermine the consequentialist's plausible insight: that it is always permissible to do what is best (Hurley 2013; Muñoz 2021). The compellingness of this insight drove Philippa Foot to exclaim:

> It is remarkable how utilitarianism tends to haunt even those of us who will not believe it. It is as if we for ever feel that it might be right, although we insist that it is wrong.   (Foot 1983, p. 273)

After all, how could it be wrong to do what is best? But once we understand what the contemporary consequentializer means by doing what is best, the consequentialist's insight is no longer so compelling. For example, suppose I claim that it is always right to do what is most beneficial to American bullfrogs. I might even claim that the goodness of outcomes is entirely determined by the benefits provided to American bullfrogs. But now the consequentialist's compelling insight becomes the claim that it is always permissible to do what is most beneficial to American bullfrogs. This insight is unlikely to haunt or compel us.

My own program has more in common with the Good Old Days than with newer consequentializing approaches. As a result, these objections will not apply to my project. My views will certainly be nontrivial: at many points, I will claim that my opponents recommend irrational patterns of inquiry. I will do my best to theorize about the ordinary notion of the good. I will also retain the consequentialist's traditional explanatory ambitions. Here too, I will argue at chapter length (Chapter 6) that my account delivers plausible, correct, and unifying explanations of disparate phenomena about rational inquiry. And although the matter is not for me to judge, I hope to do all of this in a way that preserves the consequentialist's compelling insight that it cannot be wrong to do what is best.

3.6 Conclusion

My project in this chapter was to develop a normative theory of rational inquiry for bounded agents. The account has three main components.

First, I adopt a global consequentialist account of the norms governing features of agency. On this account, features of agency are right just in case they are best, and the value of features of agency is determined by the value that they promote. The reason-responsive consequentialist view is a version of expansionist global consequentialism insofar as it aims to extend this account of right agency to cover other normative terms beyond rightness. In particular, I aim to develop a global consequentialist account of rational agency.

The second feature of the RRCV shows how to extend global consequentialism from rightness to rationality. Here I adopt a reason-responsiveness account of rationality. On this account, rationality consists in responding correctly to possessed normative reasons. I use this account to recover a global consequentialist account of rationality, on which rationality is a strictly more demanding status than rightness. Rational agency requires not only doing what we ought but also responding correctly to the normative reasons that we possess for doing so.

From what perspective are the reasons that we possess determined? The third component of the RRCV answers this question using an information-sensitive reading of deontic modals. On this account, the reasons that we possess are determined not by the totality of facts, nor by our beliefs about the world, but rather by a relevant body of information such as an agent's total evidence. I argue that an information-sensitive account of deontic modals improves on objectivism by incorporating agents' informational bounds, and improves on subjectivism by accommodating agents' limited abilities to incorporate relevant information into their beliefs.

I also introduced three subsidiary components of the reason-responsive consequentialist view. The first is a sharp level separation between normative questions about belief and inquiry. This will allow us to hold that traditionally non-epistemic reasons are robustly relevant to rational inquiry without immediately implying that the same is true of rational belief. The second component is a rich axiology on which many achievements, both intellectual and non-intellectual, bear final value. This will allow us to recover a broad range of truth-directed norms as well as to respect the intuition that intellectual achievements are valuable in their own right. Both of these components form part of a third component of any good consequentialist account: a consequentializing program. Consequentializing aims to recover unifying, nontrivial, detailed, and plausible explanations of normative phenomena in consequentialist fashion.

This completes the development of my positive theory. In Part 3, I defend the RRCV in its application to rational inquiry, and in Part 4 I apply the RRCV to understand the meaning of key normative claims about bounded rationality as well as to assess their correctness and implications.

Before proceeding, I want to address an objection that is often raised to my account. The objection is that my account changes the subject. The question at hand is what epistemic rationality demands of us, but the RRCV is not obviously an account of epistemic rationality. Does the account succeed by changing the subject? This is an important question, and it deserves a straight answer. I take up this question in Chapter 4.

4

There are no epistemic norms of inquiry

4.1 Introduction

Chapter 3 developed a reason-responsive consequentialist view of rational inquiry. On that view, acting rationally consists in doing what we have most reason to do in response to the reasons for which we ought to do it. What we have most reason to do is to promote value, and the promotion of value is assessed in an information-sensitive way. Since inquiry is an activity, an account of rational inquiry falls out of the RRCV as a special case.

An immediate objection to the reason-responsive consequentialist view is that it answers the wrong question. As a consequentialist view, the RRCV does not aim to characterize a special type of epistemic rationality, but rather rationality full stop. Is this a problem for the RRCV?

One way of responding to the objection would be to concede the existence of epistemic norms governing inquiry and to push back against *epistemic prioritarianism* for inquiry: the claim that epistemic norms governing inquiry deserve special emphasis in epistemological theorizing. I have taken this approach elsewhere (Thorstad 2021). The rejection of epistemic prioritarianism would be enough to motivate my project, and I will discuss at least one position (Section 4.7.3) which may favor this line of response. Readers who are prepared to reject epistemic prioritarianism for inquiry are welcome to stop reading and continue on to Part 3.

In this chapter, I want to explore a stronger line of response: there are no epistemic norms of inquiry. Call this view *epistemic nihilism for inquiry*. If epistemic nihilism for inquiry is true, then it is no objection to the RRCV that it fails to characterize the epistemic norms governing inquiry, because there are no such norms.[1]

4.2 Strategy and motivation

Epistemic nihilism for inquiry is a strong view, so it may help to begin with three preliminary motivations. First, until recently epistemic nihilism was the received stance toward inquiry. Many epistemologists claimed that epistemic norms govern

---

[1] This chapter is adapted from Thorstad (2022a).

doxastic attitudes rather than the inquiries that produce them, and some defended the further claim that epistemic norms are purely synchronic (Doughtery 2014; Feldman 2002; Hedden 2015). While these views are no longer as popular as they once were, it is important to make sure that they get their day in court.

Second, many recently proposed norms of inquiry, or *zetetic norms*, do not look epistemic. These norms make straightforward appeal to traditionally non-epistemic factors such as an agent's interests and goals. For example, both classic (Harman 1986) and modern (Friedman 2018) formulations of clutter avoidance understand it as a norm against inquiring into matters unrelated to the inquirer's interests and desires. And Jane Friedman has defended a number of explicitly instrumentalist norms, such as the following:

> **(Zetetic Instrumental Principle (ZIP))** If one wants to figure out $Q^?$, then one ought to take the necessary means to figuring out $Q^?$.   (Friedman 2020, p. 503)

In this vein, many zetetic epistemologists have claimed that rational inquiry has a significant practical component (Friedman 2020; Harman 2004; Lord 2020), and some have questioned whether epistemic normativity provides the most interesting lens into rational inquiry (Thorstad 2021). A natural way to make sense of these trends would be to claim that zetetic norms are not, in fact, epistemic norms, but rather instrumental, practical, moral or all-things-considered norms.

Third, it has recently been claimed that plausible zetetic norms such as ZIP are in tension with traditional epistemic norms (Friedman 2019b, 2020). If zetetic norms are also epistemic norms, this tension threatens to imply that traditional epistemic norms are in need of revision. But if zetetic norms are not epistemic norms, then this tension may be no more surprising or problematic than familiar tensions between epistemic and all-things-considered norms.

These are, of course, only preliminary motivations. They are not decisive arguments for epistemic nihilism about inquiry, and there is plenty that could be said in reply. What would the strongest case for epistemic nihilism about inquiry look like?

Ideally, epistemic nihilists would begin by addressing existing arguments against epistemic nihilism for inquiry. A problem for this strategy is that few direct arguments have been given against epistemic nihilism for inquiry. So while I will address the most extended argument of which I am aware (Section 4.7.4), we may need to look elsewhere to ensure that epistemic norms are given a fair shake.

My proposal is to return to the arguments which originally led us to posit a distinctive type of epistemic normativity governing belief and ask whether these arguments also give us good grounds to posit epistemic norms governing inquiry. In each case, I show that the arguments fail to generalize as arguments against epistemic nihilism for inquiry, and that in many cases the arguments tell in favor of epistemic nihilism for inquiry. If successful, this strategy will put the burden

on defenders of epistemic norms for inquiry to produce new grounds on which epistemic norms should be posited. It may also suggest that epistemic nihilism about inquiry is supported by the best case *against* epistemic nihilism for belief, or that epistemic norms for inquiry should be regarded as an overgeneralization on the original motivations for positing epistemic norms governing belief.

Here is the plan. In Section 4.3, I make six remarks to clarify my project. Sections 4.4–4.6 survey three leading arguments against epistemic nihilism for belief: the *argument from non-existence* that there are no non-epistemic reasons for belief; the *linguistic argument* that epistemic norms are needed to account for the semantics of epistemic terms; and the *argument from theoretical roles* that epistemic norms are better-suited than all-things-considered norms to play key theoretical roles associated with rational belief. In each case, I show that these arguments fail to generalize, and in many cases tell in favor of epistemic nihilism for inquiry. Section 4.7 discusses objections. Section 4.8 develops an alternative Gibbardian picture on which norms of inquiry are all-things-considered norms governing action. Section 4.9 concludes.

4.3 Clarifying the view

Before beginning, I want to say six things by way of clarifying my target. First, epistemic nihilists about inquiry do not claim that recently proposed zetetic norms such as ZIP are false or unimportant. Quite the opposite: a problem for zetetic epistemologists is that many zetetic norms look more like instrumentalist norms than traditional epistemic norms. My aim is to explain why these norms can be true and important, even if we are disposed to deny that instrumentalist norms of belief would be true or important.

Second and relatedly, we need to distinguish the question of whether zetetic norms are epistemic norms from the question of whether zetetic epistemology is properly understood as a type of epistemology. Zetetic epistemologists have given compelling arguments that the study of zetetic norms is an important project within epistemology: rational inquiry is a central component of theoretical rationality (Kelly 2002, 2003; Thorstad 2021); throughout history, epistemologists have often been concerned with inquiry (Friedman 2017b; Misak 1987; Striker 2001); and a purely synchronic epistemology risks engaging in temporal parochialism, by which we study attitudes while ignoring the temporally extended processes that produced them (Friedman 2020). None of these arguments turns in any obvious way on the claim that zetetic norms are epistemic norms. And as Friedman (2020) has emphasized, even in the case of belief there have often been instrumentalists, pragmatists and other skeptics about epistemic normativity. We may think that these theorists are wrong about what rationality requires, but we do not often go to the extreme of denying that they are doing epistemology.

Third, we need to distinguish the question of whether there are epistemic norms governing the *activity* of inquiry itself from the further question of whether an agent's inquiries can affect the normative status of her doxastic attitudes.[2] Epistemic nihilism about inquiry deals with the first question, but says nothing about the second. For example, it has long been held that gathering evidence can increase the justification of an agent's beliefs, by providing evidence to support them. And more recently, some authors have held that failures to gather evidence can make an agent's beliefs unjustified (Baehr 2009; Miracchi 2019). Both of these claims should be distinguished from a further claim about inquiry: that we have epistemic duties to gather evidence in order to increase the justification of our beliefs. Many traditional epistemologists have denied the existence of these and other inquiry-related duties, while leaving room for the claim that inquiries can affect the normative status of beliefs (Feldman 2002). I return to this example in Section 4.

Fourth, to deny that there are epistemic norms governing inquiry is to deny that a certain class of normative vocabulary takes a distinctively epistemic reading. My focus in this chapter will be on four normative terms: rationality; justification; and what agents should or ought to do. I will argue that we lack adequate reason to posit a distinctively epistemic sense in which inquiries count as rational, justified, or as inquiries that agents should or ought to engage in. I focus on these terms because they figure in the formulations of many recent norms governing inquiry (Friedman 2020; Woodard forthcoming a) as well as in many classic debates about epistemic norms governing belief.

In this chapter, I will not be explicitly concerned with a variety of other normative terms such as virtue (Whitcomb et al. 2017), fittingness (McHugh and Way 2016), reasons (Fleisher 2022), blameworthiness (Brown 2020), pursuitworthiness (Fleisher 2022; Šešelja and Straßer 2014; Whitt 1992), objectivity (Longino 1990), bias (Gilovich and Griffin 2002; Lee and Schunn 2011), or the aim or function of inquiry (Falbo forthcoming; Friedman ms; Kelp 2021b).[3] My arguments may have direct implications for these terms if the terms are linked by bridge principles to claims about rationality or other normative categories. But if there is no especially direct route between, for example, claims about the aim of inquiry and claims about how agents ought to inquire, then my discussion may not directly pronounce on the question of whether inquiry has a distinctively epistemic aim. And in

---

[2] This discussion will become more complicated if you think that inquiry requires adopting a certain doxastic attitude: suspension of judgment (Friedman 2017a). We could avoid this complication by holding that the relevant sort of suspension is not a doxastic attitude (McGrath 2021) or by posing challenges to the requirement (Lee forthcoming; Masny 2020; Millson 2021; Palmira 2020).
[3] Thorstad (2021) suggests that we may want to shift scholarly attention toward such terms if we want to make room for claims that are not true about rationality. Perhaps the existence of distinctively epistemic norms is one claim which benefits from a shift in focus.

particular, I do not mean to deny that there are epistemic reasons for inquiry. I discuss the existence of epistemic reasons for inquiry in Section 6.3.

Fifth, my discussion in this chapter is not premised on the denial of pragmatic, instrumentalist, or other traditionally non-epistemic views about rational belief (Rinard 2019b; Steglich-Petersen and Skipper 2019). Quite the opposite: arguments for epistemic nihilism about belief may provide good support for epistemic nihilism about inquiry.[4] However, I aim to show that even if the traditional case against epistemic nihilism for belief is compelling, that case does not tell against epistemic nihilism for inquiry, and may well support it.

Finally, what does it mean to call a norm epistemic? It has recently become clear that there are many things this claim could mean (Conee 2016; Cohen 2016a, b; Lyons 2016; McGrath 2016). Some of these will be my targets in this chapter, whereas others will not.

Here are two examples of what my target might be. On a *value-based conception*, epistemic norms are picked out by a special type of epistemic value they direct us to promote, honor, or instantiate. The value-based conception is familiar from recent work in epistemic consequentialism, as well as from Richard Foley's Aristotelian conception of rationality (Foley 1987). I discuss the value-based approach in Section 6.5. On an alternative *knowledge-based conception*, epistemic norms are picked out by their close relationship to knowledge. Traditionally, epistemic justification was understood to pick out whatever plays the role of the 'J' in the JTB + X analysis of knowledge, although broader versions of the knowledge-based conception are possible.[5] I discuss a version of the knowledge-based conception in Section 5.1.

Here are some examples of what my target is not. Some theorists understand the term 'epistemic' to pick out norms governing belief, as opposed to norms governing other objects such as inquiry. On this conception, epistemic nihilism for inquiry would be trivially true. Alternatively, Jessie Munton suggests we use the term epistemic "to indicate a flaw that arises in virtue of an irrational response to information, or through a lack or loss of information" (Munton 2019, p. 231). On this conception, most or all norms of inquiry may count as epistemic, since inquiry involves gathering, storing, and responding to information, in which case epistemic nihilism for inquiry would be trivially false.

What distinguishes those conceptions of the epistemic which concern me from those that do not? I am interested in conceptions of epistemic normativity that would generate a nontrivial distinction between epistemic and non-epistemic norms of inquiry and ground arguments for dedicating a significant portion of zetetic epistemology to the study of epistemic norms governing inquiry. Because

---

[4] This is one way to read Steglich-Petersen (forthcoming).

[5] Here are two broader examples. Conee (2016, p. 858) holds that "epistemic justification is the kind of justification that pertains especially to knowledge or cognition" and Fricker (2007) holds that injustice is epistemic when it wrongs agents in their capacity as knowers.

several different conceptions of the epistemic could be at play here, my strategy will be to focus not on any particular conception of epistemic normativity, but rather on the arguments which have been traditionally given for positing a distinctive type of epistemic normativity. If I am correct that these arguments do not generalize to the case of inquiry, then more work will be needed to ground a call for the study of epistemic norms governing inquiry.

Summing up, epistemic nihilism for inquiry does not claim that zetetic norms are false, unimportant, or not a subject for epistemologists. Epistemic nihilism is only a claim about the norms governing inquiry, not about the normative status of beliefs that result from inquiry. Epistemic nihilism for inquiry takes no stance on the correctness of epistemic nihilism for belief. In this chapter, I focus on the prospects for epistemic nihilism about four normative terms: justification, rationality, and how agents should or ought to inquire. I target any conception of epistemic normativity that would generate a nontrivial distinction between epistemic and non-epistemic norms of inquiry, grounding arguments for dedicating a significant portion of zetetic epistemology to the study of epistemic norms governing inquiry.

With these clarifications in mind, I survey three motivations that could be given for rejecting epistemic nihilism about belief. In each case, I argue that these motivations may well be compelling against epistemic nihilism for belief, but cannot tell against epistemic nihilism for inquiry and in many cases, may even support it.

4.4 The argument from non-existence

One of the most common arguments against epistemic nihilism for belief is that there are no non-epistemic reasons for belief. In this case, it is natural to conclude that belief is governed by a distinct type of epistemic normativity which answers only to epistemic reasons. Call this the *argument from non-existence*. We might try to extend the argument from non-existence to show that there are no non-epistemic reasons for inquiry. This extended argument from non-existence would tell against epistemic nihilism for inquiry.[6]

---

[6] One interesting question which we will return to in Section 4.7 is whether even the view that all reasons for belief (or inquiry) are epistemic need force us to posit a distinctive type of epistemic rationality governing belief (or inquiry). One reason for skepticism on this front is that, by hypothesis, the posited notion of epistemic rationality would coincide with an all-things-considered notion (Section 4.8), so unless we deny the cogency of all-things-considered rationality for belief (or inquiry), epistemic rationality would seem on this view to be a dispensable posit. We could, of course, rephrase Shah's point about transparency as a point about conceptual roles (what settles belief) rather than reasons, but this would collapse into the subpart of the argument from theoretical roles discussed in Section 4.6.2 and so it would not be a distinct argument against epistemic nihilism. Thanks to a referee for pushing me to address this point.

In this section, I survey three leading arguments against the existence of non-epistemic reasons for belief and ask whether any of them tells against the existence of non-epistemic reasons for inquiry. I show that each of these arguments not only fails to support the conclusion that there are no non-epistemic reasons for inquiry, but in fact supports the existence of non-epistemic reasons for inquiry.

A traditional motivation for the claim that all reasons for belief are epistemic is due to Thomas Kelly (2002, 2003). Normative reasons for belief should be potential motivating reasons, reasons on which our beliefs can be based. But Kelly argues that non-epistemic reasons cannot be motivating reasons for belief, and hence cannot be normative reasons for belief.[7] For example, the desire to go to heaven can be a reason to gather and attend to evidence for God's existence. This desire can also be a reason to adopt belief-like attitudes such as acceptance or faith. But I cannot believe that God exists on the basis of my desire to go to heaven. That belief must be based on my assessment of the available evidence.[8] On this basis, Kelly argues that defenders of non-epistemic reasons for belief commit the *consequentialist mistake* of assuming that since the consequences of actions bear on their rationality, the consequences of beliefs bear on their rationality as well. In doing so, we ignore the fact that consequences can be motivating reasons for action, but not for belief.

This argument poses no threat to the existence of non-epistemic reasons for inquiry because inquiry is an activity.[9] Non-epistemic reasons can be, and frequently are motivating reasons for inquiry. We can inquire about the weather in order to gain knowledge (Kelp 2021a, b) or assuage our curiosity (Whitcomb 2010), but also to plan a picnic. In this way, denying the existence of non-epistemic reasons for inquiry commits the *reverse-consequentialist mistake* of assuming that since the consequences of beliefs do not bear on their rationality, the consequences of inquiries which produce belief do not bear on the rationality of inquiry (Thorstad 2021).

Kelly himself is quite friendly to this conclusion. Kelly holds that theoretical rationality is a hybrid virtue. While the rationality of belief is a purely epistemic matter, theoretical rationality encompasses both belief and inquiry. A lively mixture of epistemic and non-epistemic reasons are relevant to rational inquiry. In asking which questions to inquire about, whether to gather evidence, or how

---

[7] Relatedly, one might argue that non-epistemic reasons can only motivate voluntary features of agency, but that beliefs are not voluntary in the relevant sense. Again, this argument would not generalize to inquiry, which is a voluntary act. Thanks to an anonymous referee for raising this important point.

[8] For pushback, see Leite (2007), Rinard (2015), and McCormick (2015).

[9] On some views, inquiry is coextensive with the possession of interrogative attitudes (Friedman 2017b). But this does not rule out the existence of practical reasons for inquiry unless we deny the existence of practical reasons for interrogative attitudes. This move looks particularly implausible once we recall that for Friedman, many interrogative attitudes are actions rather than attitudes (Friedman 2013, 2017b).

much of the available evidence to use during reasoning, we draw not only on epistemic considerations about truth and knowledge but also on non-epistemic considerations such as the importance of answering questions and the cognitive resources consumed during inquiry. To take a purely epistemic perspective toward theoretical rationality would be a mistake, because the epistemic stance ignores a variety of systematic, difference-making reasons that rational agents respond to during inquiry.

A second motivation for the claim that there are no non-epistemic reasons for belief is due to Nishi Shah (2003, 2006). Shah's argument begins with a deliberative constraint on reasons:

(Deliberative Constraint) R is a reason for X to ϕ only if R is capable of disposing X to ϕ in the way characteristic of R's functioning as a premise in deliberation whether to ϕ.   (Shah 2006, p. 485)

The second step in Shah's argument is to appeal to the *transparency* of belief:

(Transparency) The deliberative question *whether to believe that p* inevitably gives way to the factual question *whether p*.   (Shah 2006, p. 481)

Together, the deliberative constraint and transparency imply that R is a reason for X to believe that p only if R is capable of disposing X toward believing that p in the way characteristic of R's functioning as a premise in doxastic deliberation. And, Shah argues, only epistemic considerations dispose us toward belief in this way.

Transparency poses no threat to the existence of non-epistemic reasons for inquiry. That is because the questions of whether and how to inquire into p are not transparent to whether p.[10] Perhaps there are some epistemic statuses toward p which are descriptively or normatively incompatible with inquiry into whether p, such as belief or knowledge that p (Friedman 2017a, b). But in the typical case where we neither know nor believe p, the deliberative questions of whether and how to inquire into p are answered not only by our current evidence about p, but also by the importance of the question, the cognitive costs of answering it, and the other inquiries we could engage in instead.

In fact, if we think that normative reasons are the types of considerations that settle deliberation we will be hard-pressed to avoid admitting a large and systematic class of non-epistemic reasons for inquiry. The reason for this is that it is not clear if epistemic reasons alone typically, or even ever settle deliberation about whether or how to inquire. I make this case in Section 4.6.2.

A third motivation for denying the existence of non-epistemic reasons for belief is the relocation strategy (Way 2012). This strategy redescribes purported

---

[10] This is not to deny that the questions of whether and how to inquire into p might be transparent to some other question, but only to suggest that this further question would not be settled by epistemic reasons alone.

non-epistemic reasons for belief as epistemic reasons for some activity such as getting ourselves to believe. For example, it is held that the importance of a comfortable retirement is not a reason to hold many accurate beliefs about retirement. The importance of a comfortable retirement is a reason to get ourselves to hold many accurate beliefs about retirement.

But how do we get ourselves to hold accurate beliefs about retirement? Sometimes we can shift our beliefs by attending church or popping belief pills. But the typical route to getting ourselves to have beliefs is through inquiry. If you want to form many accurate beliefs about retirement you should read books, take classes, and consult your financial planner. This means that relocated non-epistemic reasons for belief will often turn out to be non-epistemic reasons for inquiry. In this way, the relocation strategy protects against the existence of non-epistemic reasons for belief by relocating large classes of non-epistemic reasons from belief to inquiry.

In this section, we considered the argument from non-existence, which holds that we must posit epistemic norms governing belief because there are no non-epistemic reasons for belief. We asked whether the argument from non-existence could be extended to establish that there are no non-epistemic reasons for inquiry. We considered three strategies for pressing the argument from non-existence in the case of belief. We saw that, by contrast to the case of belief, each of these strategies grounds an argument for the existence of non-epistemic reasons governing inquiry. As a result, the case against epistemic nihilism for inquiry cannot be made on the grounds that there are no non-epistemic reasons for inquiry. In the next section, I consider a second argument against epistemic nihilism: the linguistic argument.

4.5 The linguistic argument

Roderick Firth (1956, 1959) introduced the notion of epistemic rationality in order to account for our ordinary epistemic talk.[11] Chisholm (1956) had proposed to account for epistemic talk using the familiar ethical notion of ought. Firth claimed that we could get a better analysis by positing a novel type of epistemic rationality, and concluded that we should posit such a notion. Call this the *linguistic argument*.

As before, the linguistic argument does give some traction against epistemic nihilism for belief. Let p be any proposition that is supported by my total evidence, but which I have significant (apparent) non-epistemic reason against believing. Perhaps p is the claim that I am a bad dancer or that my son is guilty of a crime. On the standard story, we have:

---

[11] Firth and Chisholm regarded their project as a matter of conceptual analysis. We moderns often have the weaker project of accounting for the semantics of epistemic terms, and so I have softened my presentation of the linguistic argument to match this weaker project.

(**Positive datum for belief**) The folk say that my belief in p is rational, justified, and should or ought to be held.

By contrast, suppose I believe the opposite claim: that I am a good dancer or that my son is innocent. We have:

(**Negative datum for belief**) The folk say that my belief in $\neg p$ is irrational, unjustified, and should or ought not to be held.

What should we make of these data?

The linguistic argument takes the positive datum to suggest that our ordinary epistemic talk about belief is best explained by positing a distinctive epistemic type of rationality, justification or ought that applies to belief. The linguistic argument also suggests that this epistemic reading is important, insofar as it figures in much of our everyday thought and talk about belief. By contrast, the linguistic argument takes the negative datum to put pressure against the existence or importance of all-things-considered norms governing belief. If there were a distinctive type of all-things-considered ought, rationality or justification which applies to belief, we would expect the folk to say so. If folk discourse does not track all-things-considered norms, then that constitutes some evidence that these norms do not exist, or at least that they are not at issue in much of our ordinary thought and talk about belief.

But the situation with inquiry is precisely the reverse. Let I be an inquiry which I have most all-things-considered reason to undertake, but not most epistemic reason to undertake. Examples will vary depending on what we understand epistemic reasons for inquiry to be, but on many views it might be helpful to think of I as an inquiry that is practically important but promises little in the way of novel truth, knowledge, or understanding. In this case, we seem to have:

(**Positive datum for inquiry**) The folk say that performing I would be rational, justified and should or ought to be done.

(**Negative datum for inquiry**) The folk say that any alternative I' to I would be irrational, unjustified and should or ought not to be done.

In a moment, I will provide evidence for the positive and negative data for inquiry. But first, consider what these data would imply.

A parallel linguistic argument would take the positive datum for inquiry to suggest that our ordinary epistemic talk about inquiry is best explained using all-things-considered readings of rationality, justification, and ought. This linguistic argument would also suggest that the all-things-considered readings are important, insofar as they figure in much of our everyday thought and talk about inquiry. By contrast, a parallel linguistic argument would take the negative

datum to put pressure against the existence or importance of epistemic norms governing inquiry. If there were a distinctive type of epistemic ought, rationality or justification which applies to inquiry, we would expect the folk to say so. If folk discourse does not track epistemic norms, then that constitutes some evidence that these norms do not exist, or at least that they are not at issue in much of our ordinary thought and talk about belief.

This means that if the positive and negative data for inquiry are on the right track, the linguistic argument cannot be taken to support the existence and importance of epistemic norms governing belief without at the same time telling against the existence and importance of epistemic norms governing inquiry. But why should we believe the inquiry data?

One way to motivate the inquiry data is to look at cases. Consider:

(Smoke) You are reading a stimulating philosophical book when you notice smoke outside your window. You must choose whether to continue reading or to pause your philosophical inquiry and investigate the smoke.

We can imagine ways of filling out the case in which you have most epistemic reason to investigate the smoke, but also continuations in which you have more epistemic reason to continue reading. The point is that folk judgments in this and similar cases are entirely insensitive to variations in epistemic reasons unless these shifts are sufficient to change what you have most overall reason to do. Consistent with the positive datum, the folk say without hesitation that you should investigate the smoke, and would be rational and justified in doing so. Consistent with the negative datum, the folk say that you should not continue reading, and would be irrational and unjustified in doing so. And the folk do not change their judgments until we shift the all-things-considered balance of reasons for inquiry.

Another way to motivate the inquiry data is to look at proposed epistemic norms governing inquiry. Many proposed epistemic norms sometimes permit doing what we do not have most all-things-considered reason to do or forbid doing what we have most reason to do. In these cases, the inquiry data predict that folk judgments will come apart from the mandates of proposed epistemic norms. And indeed, that seems to be the case.

For example, consider one of the best-known epistemic norms on evidence gathering:

(HJ) For any proposition that is less than certain on one's present evidence, one has an epistemic duty to seek more evidence about that proposition.

(Hall and Johnson 1998, p. 133)

Now consider a detective who stops gathering evidence each night so she can spend time with her family when she could have instead continued inquiring into

some uncertain proposition. Consistent with the positive datum and against HJ, we say without hesitation that she ought to go home rather than continue gathering evidence and that going home is justified, and rational. And as the negative datum predicts, if she were to stay home late to gather evidence we would say that she had acted as she ought not and that her actions were unjustified and irrational. These folk judgments are unswayed by the fact that the detective could have gathered more evidence by working harder, so long as the detective had more reason to return home than to continue gathering evidence.

So far, we have seen that the positive and negative data for inquiry can be supported by reflection on cases, as well as by considering situations in which proposed epistemic norms governing inquiry come apart from all-things-considered norms. If this discussion is on the right track, then the very same linguistic data used to support the existence and importance of epistemic norms governing belief should, if anything, be taken to tell against the existence and importance of epistemic norms governing inquiry. This means that the linguistic argument cannot refute epistemic nihilism for inquiry, and may tell in favor of nihilism.

In the next section, I consider a final reason for positing epistemic norms: that they are the best candidates to play important theoretical roles attached to the concept of rationality. Again, I argue that these data at best provide no evidence against nihilism, and in some cases may support nihilism.

4.6 Theoretical roles

A third argument against epistemic nihilism for belief is that the notion of epistemic rationality is needed to play key theoretical roles that we would like the notion of rational belief to play. These include figuring in the conceptual analysis of knowledge (Section 4.6.1) and settling doxastic deliberation (Section 4.6.2). Although these may be good motivations for positing epistemic norms governing belief, I argue that they do not generalize to motivate epistemic norms for inquiry, and in fact that the second role is best played by all-things-considered norms rather than epistemic norms for inquiry.

4.6.1 The analysis of knowledge

For the early analytic epistemologists who introduced and popularized the notion of epistemic justification, the notion of epistemic justification was strongly, and sometimes even constitutively identified with its role in the analysis of knowledge.[12] Epistemic justification was needed to play the role of the 'J' in the

---

[12] For a representative early view, Bonjour (1985, p. 5) holds that epistemic justification is "the sort of justification pertaining to and appropriate to knowledge." This view is echoed in recent authors such

JTB + X analysis of knowledge. Although some contemporary epistemologists, including many knowledge-firsters, have soured on this project (Kelp 2021a; Williamson 2000), a considerable fraction of epistemologists would like to make room for the traditional analysis of knowledge.

If epistemic justification is meant to play its familiar role in the analysis of knowledge, then most plausible moral, prudential or all-things-considered notions of epistemic justification will be nonstarters. The reason is that these accounts generate the wrong kinds of pragmatic encroachment on knowledge. They say that beliefs can be justified, in the sense that counts toward knowledge, because it is valuable to be in a belief state itself. For example, a Pascalian belief that God exists might be justified as a safe ticket to heaven.

Now some philosophers do accept other patterns of pragmatic encroachment on knowledge. For example, practical stakes might raise the level of evidence needed for beliefs to be candidates for knowledge (DeRose 2009; Fantl and McGrath 2009). But I know of few philosophers who think that beliefs can be candidates for knowledge on the basis of the practical or all-things-considered utility of occupying a belief state itself as a means to achieving other goals.[13] A Pascalian true belief does not become a better candidate for knowledge if it turns out that the belief greatly improves your odds of reaching heaven.

Insofar as we want the notion of epistemic justification to play its traditional role in the analysis of knowledge, many philosophers have thought that we have good reason to posit a novel notion of epistemic justification because it is the best candidate to play this role.[14] This is, for philosophers interested in the analysis of knowledge, a compelling reason to reject epistemic nihilism for belief. But this argument does not give us any direct reason to reject epistemic nihilism for inquiry, since knowledge is a status that attaches to beliefs rather than inquiries. We saw in Section 4.3 that the mere observation that inquiries can alter important epistemic statuses of belief, such as justification and knowledge, does not yet force us to posit a derivative class of epistemic duties for inquiry.

So far we have seen that there is one theoretical role, figuring in the analysis of knowledge, which tells against epistemic nihilism for belief but not against epistemic nihilism for inquiry. Next, I argue that there is at least one prominent theoretical role that tells against epistemic nihilism for belief, but *in favor* of epistemic nihilism for inquiry.

---

as Conee (2016), Greco (2011), and Goldberg (2016). And prior to 1948, the term 'epistemic' was not even associated with a type of justification: it simply meant 'of or pertaining to knowledge' (Marcucilli 2019).

[13] Here Alston (1978, p. 277) is typical: "what counts towards *S*'s knowing that *p* is not that he is morally, prudentially, or legally justified in believing that *p*, but rather that his belief that *p* satisfies some specifically epistemic standards, standards that have to do with a kind of excellence that is appropriate to the quest for knowledge."

[14] Indeed, this argument was first made by Firth (1956) himself.

4.6.2 Settling deliberation

There is a robust tradition in epistemology that links the rationality of belief to the conditions under which doxastic deliberation is properly settled (Kelly 2002; Shah 2003; Williams 1973). Here is a pared-down version of that tradition that many epistemologists would accept.

We can pose to ourselves the *deliberative question* of whether to believe that p. This question is settled by the factual question of whether p. And the factual question of whether p is settled by evidence for and against p. Now you might think that an important theoretical role for rational belief is to track the conditions under which the deliberative question of whether to believe p is properly settled. If that is right, then two consequences follow. First, we have good reason to posit a distinctively epistemic type of rationality that is responsive only to evidence, since evidence is what settles the deliberative question of whether to believe that p. Second, we have some evidence against other non-epistemic notions of rational belief, insofar as they do not track the considerations which settle the deliberative question.

This argument may well support the existence of epistemic norms for belief, but if anything it tells against the existence of epistemic norms for inquiry. We can also pose the deliberative question of whether or how to inquire into some question Q. Yet this deliberative question is not settled by the *epistemic question* of how we have most epistemic reason to inquire, but rather by the *all-things-considered question* of how we have most all-things-considered reason to inquire. There are two ways to see that the all-things-considered question, rather than the epistemic question, settles the deliberative questions of whether and how to inquire.

First, many philosophers accept a Humean theory of motivation on which agents cannot be moved to act without the presence of desire or another conative state.[15] A Humean theory predicts that the mere fact that some inquiry conduces to epistemic goods such as truth, knowledge, and understanding will be insufficient to settle the deliberative question by moving us to inquire. If we have no desires, anxieties, curiosities, or other partially conative states caught up in the inquiries at hand, we may refuse to engage in inquiries while recognizing that we have most epistemic reason to do so.[16]

Consider an example due to Kelly (2003). I approach you at the ticket counter offering cost-free evidence in the form of movie spoilers. Epistemically speaking

---

[15] There may be some tension between Transparency and Humean theories of motivation. Why should desires be needed to settle theoretical deliberation if they are not needed to settle practical deliberation? There are also other constraints to be raised about the meaning of, and motivation for, Transparency (Rinard 2017, 2019a). Concerns about Transparency can only help the epistemic nihilist, since they suggest that the case against epistemic nihilism may be weaker than supposed, even in the case of belief. Thanks to a referee for raising these points.

[16] Would matters change if we took Peirce (1877) to show that the recognition of doubt grounds a type of anxiety that pushes toward the resolution of doubt? Not obviously, for we could still hold that these anxieties are often too weak to be motivationally decisive. For illustration, see the example by Kelly in the next paragraph and also Hertwig and Engel (2021).

you have only truth to gain from accepting my offer, and truth to lose if you refuse. Yet you will likely refuse my offer because movie spoilers are truths you desire not to learn. And if you have no desire for some epistemic good such as a true belief about the ending of a movie, then learning that you could achieve this good through inquiry will not settle the deliberative question in favor of inquiring.

Second, we can see that the deliberative question for inquiry is settled in an all-things-considered manner by reflecting on intuitions about proper settling. Consider again:

> **(Smoke)** You are reading a stimulating philosophical book when you notice smoke outside your window. You must choose whether to continue reading or to pause your philosophical inquiry and investigate the smoke.

In Smoke, I submit that you both will and should put down your book to investigate the smoke. Now we can imagine ways of filling out the story on which you would learn more or weightier truths by investigating the smoke, but also continuations on which you stand to gain more truth, knowledge, understanding, and other intellectual goods by continuing to read. The point here is that our judgments about what does and should settle you in favor of investigating the smoke rather than the book in such cases are not only, or perhaps not even primarily sensitive to these epistemic questions. They track something more like the all-things-considered question of what you have most reason to do, and in this case that is dominated by practical considerations, such as the need to prevent your house from burning down.

Hence insofar as rationality is meant to track what settles the deliberative question of whether or how to inquire, it is an all-things-considered rather than an epistemic reading of rational inquiry that we are after. Together with our earlier discussion of knowledge, this suggests that thinking about theoretical roles for rationality does not clearly support the existence of epistemic norms for inquiry. Some roles, such as the analysis of knowledge, apply only to belief and not inquiry, whereas others, such as settling deliberation, may support a non-epistemic reading of norms of inquiry. Of course, the defender of epistemic norms for inquiry is always free to identify new theoretical roles which epistemic norms are needed to play. The upshot of this section is then an open invitation to describe these roles and to explain why epistemic norms are needed to play them.

4.7 Objections and replies

So far, I have defended epistemic nihilism for inquiry: the view that there are no epistemic norms of inquiry. My strategy was to examine motivations for rejecting epistemic nihilism about belief and ask whether these motivations also tell against epistemic nihilism for inquiry. I considered the argument from non-existence that there are no non-epistemic reasons for belief; the linguistic argument that an

epistemic reading is needed to capture much of our ordinary epistemic talk; and the argument from theoretical roles that an epistemic reading is needed to play key theoretical roles associated with rationality. In each case, I argued, these motivations do not generalize to support the existence of epistemic norms for inquiry, and in many cases they tell against it. Non-epistemic reasons for belief exist. Many of the same locutions that naturally take an epistemic reading when applied to belief take a natural non-epistemic reading when applied to inquiry. And the theoretical roles used to motivate an epistemic reading of rational belief either do not apply to inquiry, or motivate an all-things-considered notion of rational inquiry. These findings together put pressure on the existence of epistemic norms for inquiry and suggest that epistemic norms for inquiry may be an overgeneralization on what are otherwise strong motivations for positing epistemic norms of belief.

In this section, I consider and respond to five objections that can be raised to epistemic nihilism for inquiry.

4.7.1 All norms of inquiry are epistemic

The epistemic nihilist holds that no norms of inquiry are epistemic norms. Jane Friedman has recently defended a *unity view* on which all norms of inquiry are epistemic, including seemingly pragmatic or instrumental norms (Friedman 2020). If the upshot of Sections 4.4–4.6 is simply that we have no grounds to posit a distinction between epistemic and non-epistemic norms of inquiry, then why not take this to favor the unity view over nihilism?

I don't want to exaggerate the difference between nihilism and the unity view. In particular, nihilists are motivated by many of the same thoughts as unity theorists: a desire to make room for recently proposed norms of inquiry within epistemology; the need to avoid a temporally parochial epistemology of the present moment; and a recognition of the practical importance of inquiry within our lives. At the same time, I think that nihilism has three advantages over the unity view.

First, several theorists have raised problems for the unity view and questioned whether these data are sufficient to motivate it (Falbo ms; Steglich-Petersen forthcoming; Thorstad 2021). It can be fruitful to explore alternative views which are not subject to the same objections.

Second, the unity view is revisionary in its broad conception of what counts as an epistemic norm. The unity view asks us to accept that pragmatic, instrumentalist and all-things-considered norms can be epistemic norms. By contrast, epistemic nihilism for inquiry allows us to maintain the traditional separation between epistemic, instrumentalist, and all-things-considered norms. Blurring the lines between these classes of norms can have downstream revisionary implications, since many of the motivations Friedman cites for the unity view could equally

well be cited by discontented instrumentalists theorizing about belief.[17] Indeed, Friedman thinks there is a good chance that her view will refute evidentialism and other popular norms governing belief (Friedman 2019b, 2020).

Finally, nihilism says something stronger than the unity view: purported epistemic norms of inquiry like HJ are false. By contrast, the unity view makes room for the weaker claim that traditional epistemic norms of inquiry are true alongside traditionally non-epistemic norms of inquiry such as ZIP. If the nihilist is right that we have no strong motivation for positing such norms and some motivation for not positing them, then we should avoid positing new normative claims without necessity and adopt nihilism in favor of the unity view.

4.7.2 A terminological matter?

In conversation, it is often put to me that epistemic nihilism about inquiry is a mere terminological matter. We can use the word 'epistemic' to pick out narrow classes of norms, focused on truth, knowledge, or evidence. But we can also use it to pick out broad classes of norms, including instrumentalist or all-things-considered norms. And there is really nothing to argue about here since we can use terms however we see fit.

Now I am not sure if our dispute is terminological. I do not see myself as proposing an expanded conception of epistemic norms, but rather denying that there are any epistemic norms governing inquiry. But even if you disagree, this dispute is certainly not a *mere* terminological dispute in any sense that would render the debate uninteresting or unimportant. Since the mid-1950s, the notion of epistemic normativity has structured research in epistemology, exerting a strong influence on the questions considered to be worth addressing as well as the types of normative considerations that can be brought to bear on them.[18] It is of primary importance that we settle on the most interesting, joint-carving normative questions to ask, or else we risk going wrong by asking the wrong questions.

Consider, for example, Kristie Dotson's (2018, 2019) critique of normative epistemology. Dotson has urged that our prevailing epistemic practices serve to suppress black rage at state-sanctioned violence, casting skepticism about the official narrative as unjustified and unreasonable (Dotson 2018). As theorists, we could react to such a criticism in two ways. On the one hand, we could take it to put pressure on existing accounts of rational belief and inquiry. On the other

---

[17] Indeed, this discussion has already been used to motivate an instrumentalist unification of belief and inquiry (Steglich-Petersen forthcoming).

[18] So for example, a recent paper by David Christensen characterizes the traditional view as follows: "Epistemic rationality or justification . . . has been seen as in some sense aimed at truth, and the proper subject for epistemology. Pragmatic justification or rationality . . . has been seen as outside the purview of epistemology" (Christensen 2021, p. 501).

hand, we could dig in our heels and reply that, while interesting, the tendency of epistemic norms to countenance state-sanctioned violence is simply not the sort of thing that could count for or against a system of epistemic norms. The result of this second strategy would be that Dotson's critique, even if true, can be entirely ignored by epistemologists.

However you stand on this debate, I hope it is clear that the debate between Dotson and her opponents is not a mere terminological matter in any sense that would make the debate uninteresting and unimportant. It is of paramount importance that we decide the sorts of normative considerations that we take to be relevant to the most central and joint-carving questions in epistemology. And whatever we say about the norms governing belief, I must confess that in the case of rational inquiry I feel significant pressure to side with Dotson. For example, Dotson (2018) urges that many of us have a strong duty to gather evidence about the nature and causes of systemic injustice in order to better contextualize black rage against state-sanctioned violence. I think this claim is quite correct, and that it is important to develop and prioritize notions of rational inquiry that will ground such claims.

4.7.3 Reasons

My aim in this chapter was to show that there is no distinctive sense in which we should or ought to undertake certain inquiries, nor in which those inquiries count as rational or justified. But I have not denied that there are epistemic reasons for inquiry. The fact that an inquiry would promote truth, knowledge or understanding is certainly a reason to engage in that inquiry, just as the fact that the inquiry promotes any other valuable quantity would be a reason to engage in this inquiry. If we like, we can pick out some or all intellectual ends and reserve the term 'epistemic reason' for the fact that an inquiry would promote one or more of these ends.

There is no truth or falsity in naming, so long as referents are clearly specified. We might reserve the term 'Leonic reasons' to pick out the fact that an inquiry would benefit my dog Leo, or use 'xylophonic reasons' to pick out all reasons whose shortest perspicuous English statement begins with the letter 'x.' But precisely because there is no truth or falsity in naming, there is not much to be gained by arguing whether we should like to call some reasons 'epistemic,' 'Leonic,' or 'xylophonic.' So long as we communicate well, we may call reasons whatever we like to call them.

Perhaps one might raise the following objection.[19] In some cases, an agent's epistemic reasons will be decisive. Then it will turn out that the agent is rationally

---

[19] Thanks to an anonymous referee for raising this objection.

required to carry out the inquiry she has most epistemic reason to carry out. In this case, would it not be apt to say that there is a special epistemic type of rational obligation which the agent is under? Perhaps so, but we can also describe the situation without positing a new type of epistemic rationality. We could say, for example, that the agent is rationally required to do what she has most epistemic reason to do. Indeed, we could also say that agents are sometimes rationally required to do what they have most xylophonic reason to do; what they have most Leonic reason to do; or what their boss tells them to do. But these facts alone do not force us to posit a new epistemic reading of rationality, any more than they force us to posit xylophonic, Leonic or bossitronic senses of rationality. Epistemic rationality, like xylophonic rationality, will need to be motivated on independent grounds.

One way to go would be to hold that to any set R of reasons, there corresponds a distinctive type of R-rationality. Agents are R-rationally required to do what they have most R-type reason to do, and R-rationally permitted to do what they have sufficient R-type reason to do. Taking R to be the epistemic reasons grounds a type of epistemic rationality.

A special case of this view would be a value-based approach on which reasons are identified by the type of value that they promote, honor, instantiate, or otherwise appropriately relate to. I consider this approach in Section 4.7.5. In that section, I argue that the approach requires additional motivation; that it over-generates norms; and that for this reason it generates at most a thin sense of rationality which is not what most epistemologists have been after. I would respond in the same way to non-value-based versions of the same approach.

4.7.4 Grounds for positing epistemic norms

There is, to my knowledge, one extended argument in the recent literature for positing epistemic norms governing inquiry. This argument is due to Friedman (2020), in her defense of the unity view. That argument has a negative component and a positive component.

We have already met the negative component of Friedman's argument, which cites factors such as the need to avoid a temporally parochial epistemology of the present moment and the need to make room for instrumentalists, pragmatists and other non-evidentialist approaches within epistemology. As we have seen, these motivations are compatible with epistemic nihilism for inquiry. After all, the epistemic nihilist also thinks that epistemologists should study norms governing the temporally extended process of inquiry, and the nihilist makes room for non-evidentialist approaches to the study of zetetic norms. She simply denies that zetetic norms are epistemic norms.

On my best reading, the positive component of Friedman's argument gives three reasons why zetetic norms should be regarded as epistemic norms:

(**Govern Inquiry**) Zetetic norms govern the process of inquiry.

(**Rational Pursuit**) Zetetic norms are norms that rational subjects in pursuit of knowledge, understanding, or comprehension will follow.

(**Promotion**) Conformity to zetetic norms promotes the acquisition of knowledge, understanding or comprehension.

This argument is made in three passages (Friedman 2020, p. 505, p. 511, pp. 526–7), focusing on the Zetetic Instrumental Principle (ZIP) introduced in Section 4.2. I have reproduced and annotated one such passage below to support my reading of Friedman's argument.

> I think that ZIP has a good claim to being an epistemic norm . . . (**Govern Inquiry**) ZIP is a norm that speaks to how we should inquire. (**Promotion**) Whether or not we conform to ZIP is highly relevant to whether or not we come to know what we want to know and whether or not we succeed in understanding what we want to understand, and not merely in some superficial sense. (**Rational Pursuit**) ZIP tells us how to proceed when we want to come to know or understand something. This counts in favor of thinking of ZIP as epistemic. ZIP is a norm that a rational subject trying to know more and understand better will conform to. (Friedman 2020, p. 511)

If this is the correct reading of Friedman's argument, then how should an epistemic nihilist respond?

I don't think that Govern Inquiry will be convincing to those not already persuaded that zetetic norms are epistemic norms. Rational Pursuit would certainly be persuasive, but we need an independent argument for Rational Pursuit. On the most natural reading, Friedman accepts Rational Pursuit based on Promotion: rational agents in pursuit of intellectually valuable states such as knowledge and understanding will follow zetetic norms because conformity to zetetic norms promotes intellectual value.[20]

Now while some epistemologists do take claims such as Promotion to be decisive evidence that a norm is epistemic, many others have wanted to push back here. Consider:

(**Sandwich Norm**) If you are inquiring for many hours, you ought to pause and eat a sandwich.

We might defend Sandwich Norm's status as an epistemic norm by citing Promotion: eating a sandwich will help you to efficiently pursue truth and knowledge.

---

[20] This reading dovetails nicely with (Friedman 2019b).

But many epistemologists have thought that this causal fact alone is insufficient to make Sandwich Norm a genuine epistemic norm.[21]

What is it issue in this discussion is whether the fact that engaging in some activity would promote epistemic value grounds an epistemic requirement to engage in that activity. As it happens, there is one approach to epistemic normativity which says exactly this. I discuss this approach below since it can ground an independently motivated objection to epistemic nihilism for inquiry.

4.7.5 Existence is cheap, importance is expensive

A final objection is that on some ways of thinking about our normative obligations, the bare existence of epistemic norms is cheap. For example, consider a strong value-based approach to normativity on which for any value V and any activity A which can promote V, we have a V-type duty to engage in A in a way that best promotes V.[22] On this account, we have epistemic duties to inquire in the ways that best promote epistemic value. We also have epistemic duties to sumo wrestle in the ways that best promote epistemic value, gustatory duties to inquire in the ways that best conduce to delectable culinary experiences, and epistemic duties to pause inquiry and eat a less delectable sandwich. If this is right, then it makes little sense to deny that there are epistemic norms of inquiry so long as we can identify a distinctively epistemic type of value that inquiry promotes.

A risk of this strategy is that it threatens to over-generate duties. As we have seen, it generates not only epistemic duties to inquire, but also gustatory duties to inquire and epistemic duties to sumo wrestle and eat sandwiches. Even if there is, perhaps, some thin sense in which these duties exist and are on a par, this is not the sense that epistemologists have usually been interested in.

When we initially posited epistemic duties governing belief, we argued that these duties had important work to do. We argued that there must be epistemic duties for belief because there are no non-epistemic reasons for belief; that epistemic duties are needed to explain key features of our normative thought and talk; and that epistemic duties play key theoretical roles such as settling inquiry and figuring in the conceptual analysis of knowledge. This allowed us to claim that epistemic duties exist in a thick and highly nontrivial sense.

If epistemic duties for inquiry are to be thicker than gustatory duties for inquiry, we need to be given a story that thickens them, for example by showing how they figure in our ordinary evaluative practices or play important theoretical roles

---

[21] This discussion was popularized by Nomy Arpaly (2017), who attributes it to Sophie Horowitz.
[22] This approach could be grounded in views by Foley (1987), Friedman (2019b) and recent work on epistemic consequentialism, although many of these authors may have intended a weaker reading of the value-based approach.

associated with rationality. The purpose of Sections 4.4–4.6 is to show that the same ways in which epistemic duties for belief were thickened will not serve to put much flesh on the bones of epistemic duties for inquiry. So if thickening is called for, it must come from somewhere else.

If we do nothing to thicken the value-based notion of epistemic duties for inquiry, then this objection will not get the defender of epistemic norms of inquiry what she wants. Defenders of epistemic norms think not only that epistemic norms exist, but also that they are *important*. They want to motivate a research program that sets out and studies the epistemic norms governing inquiry, and perhaps even to put the brakes on competing all-things-considered normative programs. But so far, we have not seen any motivations for doing this, whereas by contrast we have seen several motivations for privileging the study of all-things-considered norms.

One way to read this chapter is as an invitation to defenders of epistemic norms of inquiry to thicken those norms by explaining the roles that these norms play in our thought, our talk, and our lives. But this invitation cuts both ways since I have as yet told only a partial story about the role and importance of non-epistemic norms for inquiry. It is time to remedy that. In the last section, I sketch a Gibbardian picture on which thinking how to inquire is thinking how to live, then use this picture to motivate a view on which norms of inquiry are all-things-considered norms governing action.

4.8 Thinking how to live

Inquiry is an activity. It is something we do, like building a house or writing a book. For humans, inquiry is a central activity that guides other activities by showing us how to achieve our goals. It is through inquiry that we learn to build sturdy houses and write good books. Our inquiries have profound effects on our lives and the lives of those around us. Through inquiry, we sent astronauts to the moon and became the dominant species on this planet.

Allan Gibbard taught us that thinking how to act is thinking how to live (Gibbard 2003).[23] The actions that we take determine the life that we will lead and the effects of our lives on the world around us. Nowhere is this truer than in inquiry, for our inquiries play a guiding role in almost everything that we do. Our inquiries determine whether we will obtain truth, knowledge and understanding, satisfy our curiosity and relieve doubt; but also shape whether we will be rich or poor, sick or healthy, and loved or unloved. Thinking how to inquire, like any other case of thinking how to act, involves considering all of these myriad consequences that our inquiries might have.

---

[23] To be clear, I take no stance on whether Gibbard himself would endorse the view that follows.

Consider the choices that an inquiring detective must make in investigating a murder. In *question-selection* she must decide whether to take the case, putting the question of who committed the crime onto her research agenda (Enqvist 2012; Olsson and Westlund 2006).

In *evidence gathering* she must decide how much evidence to gather, and from whom (Hall and Johnson 1998; Smith 2014). Does she interview most plausible witnesses, or only a few? Does she interview the victim's fiancé, or dismiss her as hysterical?

In *strategy selection* she must decide which reasoning strategies to use during inquiry (Lieder and Griffiths 2017; Marewski and Schooler 2011). Should she think fast, using quick and efficient heuristic rules, or think slow, using effortful non-heuristic rules (Thorstad forthcoming a, b)?

Through *inter-activity tradeoffs*, she must determine how to balance inquiry against the other activities in a rich and full life (Thorstad manuscript c). Does she spend her Friday nights working on the case, or go home and eat dinner with her family?

In *double-checking*, she must decide whether to double-check her conclusions before accusing a culprit (Friedman 2019a; Woodard forthcoming b). Might she even be required to triple-check, given the magnitude of the accusation?

In *inquiry termination*, she must decide whether to halt inquiry after leads dry up, perhaps turning the matter over to a judge or if there is not enough evidence, closing inquiry without making a final judgment and letting the case grow cold.

In all of these questions, epistemic goals such as truth and knowledge are inextricably bound up with the non-epistemic context in which the detective's inquiry occurs. To determine whether the detective should take the case, we need to know not only whether she is curious about the outcome, but also that a murder trial is at stake. To see that unusually large amounts of evidence should be gathered; stringent double-checks should be made; and unusually demanding inference rules should be applied, we need to know not only how difficult it is to identify a suspect but also what will happen if the detective's inquiry fails. In confronting inter-activity tradeoffs, we need to know not just what her inquiry will bring but also what the effects of this inquiry will be on the detective's life at home. To see that the victim's fiancé should be interviewed, we need to note not only that she is likely to have relevant information, but also that this inquiry occurs against a context in which women's testimony is often prejudicially dismissed, causing direct harm by failing to recognize women as knowers (Dotson 2011; Fricker 2007).

These remarks are not intended to marginalize traditionally epistemic considerations such as truth and knowledge, nor to suggest that rational inquiry often involves vices such as wishful thinking or slovenly reasoning meant to prevent us from discovering uncomfortable truths. If Clifford (1877) taught us anything, it is that sloppy inquiry often has devastating downstream consequences. Failing to test

the seaworthiness of a ship can lead it to sink at sea. But even as we retain Clifford's epistemic lesson that some degree of evidence-gathering is often warranted, we should not neglect the broader lesson: that an inextricable part of what warrants inquiry is the effect that inquiry will have on our lives and the lives of those around us.

If thinking how to inquire is thinking how to live, then what types of consequences bear on how we ought to inquire? All of them. We can no more neglect the effects our inquiries will have on the world than we can neglect the effects that any other action will have. If thinking how to inquire is thinking how to live, then thinking how to inquire is an all-things-considered matter. There are many ways in which our inquiries affect the world, and all of them must be considered to determine how to inquire.

We can, if we wish, tell a very different story about the norms governing belief. On this story, all reasons for belief are epistemic reasons. Seeming non-epistemic reasons for belief are really non-epistemic reasons for actions such as inquiry. Epistemic evaluations of belief states play an important role in our ordinary language, and play important theoretical roles such as figuring in the analysis of knowledge and tracking the factors which settle doxastic deliberation.

But as we have seen, this story cannot be told about inquiry, any more than it can be told for any other activity such as building a house. And that is no accident. Thinking how to inquire is thinking how to act. Thinking how to act is thinking how to live. Thinking how to live is an all-things-considered matter.

4.9   Conclusion

Chapter 3 developed an account of rational inquiry, the reason-responsive consequentialist view. A natural objection to the RRCV is that it is problematically non-epistemic. This chapter responded to the objection by defending epistemic nihilism for inquiry: the view that there are no epistemic norms for inquiry. We saw that the very same grounds which led us to posit and privilege a special type of epistemic rationality governing belief not only fail to support the existence of epistemic norms governing inquiry, but in many cases tell against the existence of epistemic norms. As a result, I suggested, we should not accept the existence of epistemic norms governing inquiry without further argument. I concluded by sketching a Gibbardian approach to inquiry on which the rationality of inquiry is an all-things-considered matter.

Having developed an account of rational inquiry and defended the account against an objection, the next task is to give positive arguments in support of the RRCV. Although Chapter 3 gave preliminary motivations for each of the RRCV's

main components, there is more to be said in support of the RRCV. Part 3 will offer two extended arguments in favor of the RRCV: the RRCV is our best hope for satisfying three minimal criteria on an account of bounded rationality (Chapter 5), and it gives precise, plausible, and unifying explanations of disparate normative phenomena (Chapter 6).

PART III
JUSTIFYING THE ACCOUNT

5

The argument from minimal criteria

5.1 Introduction

Part 2 developed an account of rational inquiry, the reason-responsive consequentialist view. Part 3 will defend the reason-responsive consequentialist view on the basis of two arguments: the argument from minimal criteria, and the explanatory argument.

This chapter develops the *argument from minimal criteria* in defense of the RRCV. The argument from minimal criteria sets out and defends three minimal criteria on an account of boundedly rational inquiry, then argues that the RRCV is our best hope for meeting those criteria. Here is the plan.

Section 5.2 introduces and defends the minimal criteria: tradeoff-sensitivity, stakes-sensitivity, and explaining the irrationality of many cases of stereotyping. Sections 5.3–5.5 consider three competing views of rational inquiry, on which inquiry aims at knowledge (Section 5.3), approximate coherence (Section 5.4), or satisfying an agent's pragmatic goals (Section 5.5). I argue that despite their merits, each view falls short by the lights of the minimal criteria. Sections 5.6–5.7 argue that the reason-responsive consequentialist view performs well by the lights of the minimal criteria. Section 5.8 concludes.

5.2 Three minimal criteria

In this section, I lay out three minimal criteria on an account of rational inquiry by bounded agents. Each criterion holds that theories of rational inquiry should be sensitive to an important type of challenge faced by bounded inquirers.

5.2.1 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. 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, we saw in Chapter 2 there is often an *accuracy-effort tradeoff* in strategy selection (Johnson and Payne 1985).[1] More cognitively demanding strategies tend to produce more accurate judgments, so agents must select strategies that 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 5.6.

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 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.

5.2.2 Stakes-sensitivity

Bounded rationality theorists have been nearly unanimous in the claim that rational inquiry is a stakes-sensitive affair. Because we cannot allocate unlimited

---

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

resources to all inquiries, we should strive, other things equal, to allocate more resources to the most important inquiries.

For example, Daniel Kahneman holds that effortful and vigilant 'system 2' cognition is often more accurate than its heuristic 'system 1' counterpart. However, system 2 cognition is also expensive, so it should be reserved for situations in which mistakes are likely or stakes are high:

> Continuous vigilance is not necessarily good, and it is certainly impractical . . . System 2 is much too slow and inefficient to serve as a substitute for System 1 in making routine decisions. The best we can do is a compromise: learn to recognize situations in which mistakes are likely and try harder to avoid significant mistakes when the stakes are high.   (Kahneman 2011, p. 44)

Similarly, while defenders of fast-and-frugal heuristics stress that heuristics can often be appropriate in high-stakes situations, most authors within this camp are well-prepared to agree with Kahneman that heuristics are more often appropriate when the stakes are low. In this vein, Laura Martignon and Kathryn Laskey write:

> Fast and frugal heuristics can have their place in everyday affairs where time is limited and knowledge is scarce, and Bayesian tools can be the choice of someone who is in no hurry and has access to a computer. Obtaining these extra percentage points of accuracy may be well worth the computational cost in high-stakes decisions such as those involving human lives or having serious long-term environmental or social impact.   (Martignon and Laskey 1999, p. 186)

Here again, it is claimed that rational inquiry is stakes-sensitive: other things equal, more cognitive resources should be devoted to higher-stakes inquiries.

To see the need for a stakes-sensitive theory, consider a case study from 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. 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.[2] Poverty creates a number of immediately pressing cognitive challenges, and agents respond by rationally reallocating the bulk of their cognitive resources toward 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.

5.2.3 Explaining the irrationality of stereotyping

We saw in Chapter 2 that a central part of the theory of bounded rationality is the claim that rational agents often make use of cognitive heuristics. But in developing this claim, we encounter a problem.

---

[2] 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).

It is an uncomfortable fact that 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 toward 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.

To see the problem in context, recall the representativeness heuristic introduced in Chapter 1. Representativeness instructs agents to judge how likely it is that an object belongs to a given category by asking how representative the object is of the category. For example, you might judge whether a distant tree is an oak by considering whether it has acorns and whether its height and shape are typical for an oak tree. In contexts such as this one, the representativeness heuristic may be a rational way to make accurate judgments at low cognitive cost.

However, we sometimes use the representativeness heuristic to make judgments about people rather than oak trees. Daniel Kahneman and Amos Tversky gave participants the following vignette:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with the issue of discrimination and social justice, and also participated in antinuclear demonstrations.
>
> (Tversky and Kahneman 1983)

Kahneman and Tversky asked one group of participants to estimate the likelihood that Linda is a bank teller, and another group to estimate the likelihood that Linda is a feminist bank teller. Participants in aggregate made the incoherent judgment that Linda is more likely to be a feminist bank-teller than a bank-teller. The explanation for this result is that agents made their judgments using the representativeness heuristic. Linda is more representative of a feminist bank-teller than a bank-teller, so participants judged that Linda was more likely to be a feminist bank-teller than a bank-teller.

Now we are in dangerous territory. In the context of social inference, the representativeness heuristic is a simple form of stereotyping. We think that Linda is unlikely to be a bank teller because she is unrepresentative of a stereotypical

bank teller. We think that Linda is an unrepresentative bank teller because she is a feminist and a woman. On the same grounds, we might judge a black man more likely to be a criminal or a Jewish man more likely to be a banker. I regard it as a non-negotiable datum that these inferences are rationally impermissible, despite their resemblance to structurally similar cases of rational heuristic inference such as representativeness-based judgments about oak trees.

A good theory of rational inquiry should say why many instances of stereo-typing can be irrational, despite their similarity to rational heuristic inferences. Otherwise, we will be forced to defend unacceptable forms of stereotyping as rationally permissible, or even rationally required, or else to reject the rationality of paradigmatically rational heuristic inferences such as representativeness-based judgments about oak trees.

Now we have our marching orders. A good theory of rational inquiry should be tradeoff-sensitive, saying how scarce resources are to be allocated within the course of a single inquiry, between inquiries, and between inquiry and other activities. It should also be stakes-sensitive, saying for example why it is generally better to invest more cognitive resources in our most important inquiries. And it should explain the irrationality of many cases of stereotyping without condemning structurally similar cases of rational heuristic inference. Sections 5.3–5.5 review three approaches to rational inquiry and argue that they do not satisfy the minimal criteria. Sections 5.6–5.7 argue that the reason-responsive consequentialist view fares well by the lights of the minimal criteria.

5.3 The knowledge aim of inquiry

Many theorists claim that inquiry aims at knowledge (Kelp 2021a, b; Millar 2011; Whitcomb 2010). In this section, I consider a popular development of that idea due to Christoph Kelp (2021a, 2021b). I argue that whatever its merits, the view does not adequately meet the minimal criteria.

For Kelp, inquiry has a constitutive aim: to settle a question. Questions are settled by coming to know their answers; hence we can also say that the constitutive aim of inquiry is knowledge. One way to ground this view is to take human activity to be divided into *critical domains* with their own constitutive aims, specifying what is valuable for its own sake in that domain. Kelp holds that inquiry is a critical domain in which knowledge is valuable for its own sake, and hence knowledge is the aim of inquiry. How does this view perform against the minimal criteria?

This view is not obviously sensitive to inter-activity tradeoffs. Suppose, for example, that inquiry is a critical domain that aims at knowledge and that house-building is a different domain that aims at building sturdy houses. Then the fact that some inquiry would take a home-builder away from building a house does not

bear on the rationality of this inquiry, since the rationality of inquiry is determined only by the aim of inquiry, which is knowledge, and not by the aim of building sturdy houses. There may be room to make space for inter-activity tradeoffs within the framework of critical domains, but I think we should say at best that this framework does not give detailed guidance about how inter-activity tradeoffs are to be made, and at worst that it threatens to block their relevance to rational inquiry.

Kelp's view might be taken to suggest that inter- and intra-inquiry tradeoffs should be made by considering what will best promote knowledge. But this view does not capture many paradigmatic claims about tradeoffs for bounded agents. For example, we would like to hold that agents should sometimes inquire using especially frugal heuristics, even if the resulting judgments will not meet the strict standards of safety or introspectability required for knowledge. We might nevertheless favor heuristics when, for example, heuristics strike the best balance on the accuracy-effort tradeoff by producing judgments that are nearly as accurate as the results of complex nonheuristic procedures, at much lower cost. But it is hard to see how the knowledge aim of inquiry can explain this. After all, neither cognitive effort, nor for that matter accuracy, figures directly in the statement of the knowledge aim.

Appeals to the knowledge aim of inquiry may be stakes-sensitive if we hold that high stakes raise the threshold for what counts as knowledge. But this would ground the wrong pattern of stakes sensitivity. If knowledge is more difficult to attain in high-stakes contexts than in low-stakes contexts, then agents who aim to acquire knowledge should be less likely to devote resources to high-stakes inquiries. And this is not the result we wanted. Other things equal, agents should be more likely, not less likely, to inquire deeply into high-stakes matters.

The knowledge aim of inquiry may explain the irrationality of many cases of stereotyping, insofar as many such inferences do not meet the demanding standards for knowledge. But the same could be said of representativeness-based inferences about oak trees. What we wanted was a view that would explain why it may be rational to apply one and the same inference procedure to categorize oak trees, but not people. We will see in Section 5.7 that there are some moves that can be made to drive these cases apart. However, there we will see that the most natural way to clarify and ground these moves rests not on brute facts about the nature of knowledge, but rather on reflection about the consequences of inquiry.

It may be possible to salvage the knowledge aim of inquiry by weakening the view. For example, if to say that inquiry aims at knowledge just means that there is something unsatisfactory or unfinished about an inquiry that fails to reach knowledge, then nothing I have said here will tell against the knowledge aim of inquiry. But by the same token, on this weaker reading of the knowledge aim, to say that inquiry aims at knowledge is not yet to say how rationality requires us to inquire. Hence we will have to look elsewhere for a complete account of rational inquiry.

In this section, we have considered a development of the view that inquiry aims at knowledge due to Kelp. We saw that this view does not adequately meet the minimal criteria. This view shows some tradeoff-sensitivity, but much of what it says about tradeoffs is difficult to square with traditional verdicts about bounded rationality. The view is either stakes-insensitive or instead exhibits the wrong pattern of stakes-sensitivity. And while the view may explain the irrationality of many instances of stereotyping, it struggles to differentiate between irrational stereotyping and rational heuristic inference in a principled way. In the next section, I consider a second view that extends the Standard Picture into an account of rational inquiry.

5.4 Approximate coherentism

What might the Standard Picture say about rational inquiry? An increasingly popular approach is to accept that bounded agents will sometimes be incoherent but to retain the normative identification of rationality with coherence. This leads to the view that bounded agents should strive to be as coherent as possible, even if full coherence is not always achievable for them (Staffel 2020; Zynda 1996). Call this view *approximate coherentism*.

Recent work has shown how to operationalize approximate coherentism in the special case of rational credence (Staffel 2015, 2017, 2020). First, identify the correct *divergence*, a generalized measure of distance, to track the distance between two credence functions. Next, measure the degree of incoherence of a credence function as the shortest distance between that credence function and a coherent credence function. By approximate coherentism, we can compare the rationality of two credence functions by comparing their degrees of incoherence: credence functions are more rational if they are less incoherent. Approximate coherentists have discussed a variety of plausible divergences for measuring credal distance and shown how many of these divergences can be used to generalize traditional arguments for the Standard Picture such as Dutch Books (Schervish et al. 2000, 2002, 2003; Staffel 2015) and accuracy-based arguments (De Bona and Staffel 2017, 2018; Staffel 2020).[3]

So far, we have seen how approximate coherentism can be spelled out as a view of rational credence. We could extend approximate coherentism to a theory of rational inquiry in many ways. For example, we could say that the rationality of inquiry processes is determined by the degree to which those processes promote actual or expected coherence. But if we try to extend approximate coherentism

---

[3] See Chapter 1 for discussion.

into an account of rational inquiry, the view we recover will struggle to meet the minimal criteria.

This version of approximate coherentism says nothing about how inter-activity tradeoffs are to be made. The approximate coherentist does not go so far as to say that we should choose between staying in the library and eating dinner with our families by asking which activity best promotes coherence. But neither does she tell us how such tradeoffs are to be made. We could, somewhat uncharitably, extend approximate coherentism to say that the rationality of non-cognitive activities such as eating dinner is determined by the amount of coherence that they promote. But this view has little appeal.

Approximate coherentism does offer advice for how inter- and intra-inquiry tradeoffs are to be made: by considering how to best promote coherence. But even when we ignore the non-cognitive stakes of inquiry, this advice remains somewhat extreme. In inquiry, bounded agents must balance competing cognitive goals such as coherence and accuracy, and these goals may come into conflict (Thorstad forthcoming). When they do, approximate coherentism says that coherence should always be privileged over other goals, such as accuracy. But insofar as goals such as accuracy are valuable, it is more natural to suppose that rational inquirers should aim to balance competing cognitive goals during inquiry, without always deciding in favor of coherence.

Approximate coherentism is not stakes-sensitive. It says that we should decide between inquiring about weekend plans and overdue electric bills by considering which inquiry would best promote coherence. That is not the result we aimed to recover. Bounded agents need to ration scarce cognitive resources with a mind to promoting not only cognitive goals such as coherence and accuracy but also noncognitive goals such as keeping the power on. A good theory of rational inquiry should be sufficiently stakes-sensitive to say how the noncognitive stakes of inquiry bear on its rationality.

Approximate coherentism will also likely fail to explain the irrationality of many cases of stereotyping. On an approximate coherentist approach, making inferences by stereotyping is irrational to the degree that these inferences tend to decrease the overall coherence of an agent's beliefs. While it may be true that those of us who generally renounce bigotry and prejudice stand to become more coherent by avoiding stereotyping, for the most committed bigots matters may be precisely the reverse. Agents who hold many prejudiced beliefs make their beliefs more, not less coherent by using those beliefs to make inferences about others. Now of course, it may turn out that even for such agents, one of the best things that can be done to improve their overall coherence is to target prejudiced beliefs and inferences wholesale. But that would be a striking coincidence in need of explanation, and I am not yet aware of any approximate coherentists who have endorsed this optimistic view.

At this point, the approximate coherentist may respond by making a distinction. She may hold that she is trying to characterize a special type of structural rationality, characterized by structural relationships between features of agency, to be carefully distinguished from substantive rationality, which tracks what agents have most reason to do (Broome 2013; Worsnip 2021).[4] Theories of structural rationality are not meant to satisfy the minimal criteria, the reply continues, so it is neither surprising nor objectionable that approximate coherentism falls short along these lines.

I am not sure if I am inclined to grant the existence of a separate type of structural rationality, although I am perfectly happy to grant that approximate coherentism is a theory of an important normative term: coherence. However, this is not the place to press that case. By now, it should be clear that my aim and the aim of many in the bounded tradition is to characterize a notion in the vicinity of what is often called substantive rationality. For my part, I would like to call the notion rationality simpliciter. But if calling my view a theory of substantive rationality is enough for the approximate coherentist and myself to understand one another, then I count it a small price to pay for successful communication.

In this section, we asked what the Standard Picture might say about rational inquiry and answered by considering an approximate coherentist approach to rational inquiry. We saw that approximate coherentism struggles to be appropriately tradeoff-sensitive, is not stakes-sensitive, and may not explain the irrationality of many cases of stereotyping. In the next section, I consider the prospects for a pragmatic account of rational inquiry.

5.5 Pragmatism

Perhaps the closest cousin of the view developed in Chapter 3 is Stephen Stich's pragmatism (Stich 1990). Stich offers the following as a 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)

---

[4] It is important to note that philosophers' distinction between substantive and structural rationality does not coincide with Simon's distinction between substantive and procedural rationality. These are separate distinctions which must be kept carefully apart.

Pragmatic views are also popular in the scientific literature on bounded rationality.[5]

Stich's pragmatism is twice removed from my own project. First, we are concerned with different evaluative focal points. Stich's view is offered as an account of how cognitive systems should be evaluated, whereas the reason-responsive consequentialist view applies to all features of agency.[6] Second, Stich and I may be concerned with different evaluative terms. The RRCV is an account of two evaluative terms: rightness and rationality. Stich's view is offered at a high level which leaves open whether it is an account of rightness, rationality, fittingness, or some other evaluative term. To speak of what holds in evaluating systems is not yet to specify which evaluative term we are interested in.[7] To precisify the view, we need to say which of these evaluative terms we are concerned with.

At the same time, I think that Stich often has a project similar to mine in mind. What would happen if we repurposed Stich's view as an account of the rationality of token inquiry processes? 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. For the moment, I will consider a version of the view on which the relevant agent is always the inquirer herself, then I will later consider what happens if this restriction is lifted.

This repurposed pragmatist view has many virtues. Among those virtues is the fact that the view meets two of the minimal criteria: it is both stakes- and tradeoff-sensitive. On this pragmatist view, agents should allocate resources between competing inquiries in the manner most conducive to obtaining what they intrinsically value. This way of settling tradeoffs is highly stakes-sensitive, with the stakes read off from the agent's own values. For example, we recover the verdict that agents are rationally required to think about how to repair a broken-down car rather than solve abstract reasoning problems, so long as they care greatly about having a working car and thinking about the car is conducive to getting it fixed. However, this pragmatist view has two drawbacks which my view avoids.

First, in many cases pragmatism does not explain the irrationality of stereotyping. On a pragmatist view, stereotyping is only irrational if the agent using the

---

[5] 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).

[6] In one sense, this means that Stich's account is less general than my own since it only deals with a single evaluative focal point, cognitive systems, and does not directly pronounce on focal points such as token processes of inquiry. In another sense, Stich's account may be more general than my own, since it is not obvious that entire cognitive systems should be understood as features of agency, in which case they may lie outside the scope of my view altogether. In fact, I tend to suspect that this must be the case, since many features of cognitive systems such as our fixed cognitive architecture are not in any sense under our control or responsive to our perceptions of normative reasons, and hence it is quite unnatural to speak of cognitive systems as things we do as agents or to ask what the right or rational cognitive system for us to have is.

[7] It does not help to talk of the system to be preferred. After all, preferences can be right, rational, fitting, virtuous, and many other things.

stereotype cares sufficiently about the welfare of others to use a more expensive inference rule instead. In the extreme case, this means that highly prejudiced agents such as our midcentury executive may be rationally required to stereotype women insofar as they are unlikely to incur social sanction for doing so, and are not greatly concerned with women's welfare. More generally, even well-meaning agents often rely too much on stereotypes. One cause of this is that although most of us are concerned for the welfare of others, we nevertheless place too much weight on our own interests and hence wrongly decide in favor of cognitive efficiency rather than justice. As a result, although pragmatism forbids most of us from engaging in severely objectionable forms of stereotyping, a pragmatic approach is likely to class many moderately objectionable cases as rationally permissible, and in fact as rationally required.

Second, although pragmatism is tradeoff-sensitive, it does not always resolve tradeoffs in a plausible direction. To see this, note that pragmatism falls out as a special case of consequentialism when we adopt an axiology on which the only things that matter are the intrinsic values of the actor at the time of action. This way of putting the view suggests a recipe for putting pressure against pragmatism. First, identify considerations that seem to matter rationally but need not be reflected in the intrinsic values of an actor at the time of action, such as the interests of others; the agent's future values; or the outcomes which hold intrinsic value but which the agent mistakenly fails to value. Next, use these gaps to construct cases in which pragmatism resolves tradeoffs in the wrong direction. For example, pragmatism holds that a selfish executive would be required to investigate a dent in her sports car rather than inquiring into the welfare of her workers if she were more concerned with the health of her sports car than with the health of her workers. Likewise, many of us recognize that our future selves will wish we resolved inter-activity tradeoffs differently, spending less time in the library and more time at home with our families. On a pragmatic view, that realization alone gives us no reason to change our present behavior, a verdict which many have found implausible (Paul 2014).

Here there are two ways that the pragmatist might respond. First, she could hold that her view characterizes a minimal Humean type of rationality which is only concerned with the satisfaction of an agent's intrinsic values. As before, I am not sure that I would like to grant this conciliatory move. However, this move has the virtue of allowing the pragmatist and consequentialist to understand each other so long as the pragmatist is willing to grant the existence of a second, non-Humean type of rationality or else to accept non-Humean views of rightness, but not rationality. However, not all Humeans are prepared to make these moves. In that case, I am happy to stake my case on the prospects for anti-Humean arguments (Korsgaard 1997; Lavin 2004).

A second pragmatist strategy would be 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 toward 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. Now we will be hard-pressed to say in any principled way which agents' interests matter to a given normative evaluation and how these interests matter. A natural way out of this difficulty is what the consequentialist proposes: to remove agents and their interests from the story altogether and talk simply about the outcomes that matter and the ways in which our actions promote, or fail to promote outcomes that matter.[8]

So far, we have introduced three criteria on an account of boundedly rational inquiry. We saw that the knowledge aim of inquiry, approximate coherentism, and pragmatism cannot adequately account for the minimal criteria. In the next two sections, I argue that the reason-responsive consequentialist view performs well against the minimal criteria.

5.6 Stakes and tradeoffs

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

Consequentialists have 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. And the stakes of inquiry reflect what actually matters, rather than what agents mistakenly take to matter. An executive who is unconcerned with her workers' welfare is still required to investigate and promote the welfare of her workers because it is a very good thing for the well-being of workers to increase.

---

[8] 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. (For suggestive remarks see (Stich 1990, pp. 134–45).) If so, this is another point at which I am comfortable parting ways.

To see that the reason-responsive consequentialist view correctly captures the normative relevance of stakes and tradeoffs, we should look toward applications where the approach 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 view 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 toward 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 reason-responsive consequentialist view 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 $\hat{x}_1, \hat{x}_2, \ldots$[9]

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.[10] On the other hand, sampling information decreases the error cost $cost(\hat{x})$ because it tends to produce a more accurate estimate.[11] Using a rich axiology, we can understand this error cost as

---

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

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

[11] 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).

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.[12] 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 my view 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.[13] 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 consequentialism to pragmatism.

More generally, there is a robust explanatory argument to be made for the RRCV based on 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 a way that explains why humans cognize as we do. 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. Making out this argument is the business of Chapter 7.

---

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

[13] Note that although increased cognitive load usually makes complex cognitive tasks more difficult, this need not always be the case (Dijksterhuis and Nordgren 2006). In such cases, it would not be rational for agents to make fewer adjustments under increased cognitive load. Thanks to a referee for pushing me to address these findings.

5.7 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 (2020). 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 my view 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.[14] 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 consequentialists need not only be concerned with the harms that will arise if the executive's 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.[15]

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. 2020), 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).

---

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

[15] This may be seen as an extension of stakes-sensitive readings of the Rule of Consideration, on which amount of value at stake combines with the probabilities of those values being promoted to determine the proper degree of consideration. Thanks to an anonymous referee for raising this suggestion.

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. And it is the same mistake that we make in judging that Linda is unlikely to be a bank teller because she is a feminist and a woman. Given the role of such inferences in creating and sustaining patterns of exclusion, marginalization and discrimination, they should be avoided in favor of more careful and individuating patterns of inference.

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 that 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. My view 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.

5.8 Conclusion

This chapter developed the argument from minimal criteria for the reason-responsive consequentialist view. Section 5.2 defended three minimal criteria on an account of boundedly rational inquiry. The right account should be tradeoff-sensitive, stakes-sensitive, and explain the irrationality of many cases of stereotyping. Sections 5.3–5.4 considered three accounts of rational inquiry: a knowledge-based view on which inquiry aims at knowledge (Section 5.3); an

extension of the Standard Picture on which inquirers aim to promote coherence (Section 5.4); and a pragmatist view on which inquiry aims to satisfy the interests of a relevant agent (Section 5.5). Although these views have many merits, we saw that each view falls short along some of the minimal criteria.

In Sections 5.6–5.7, we saw how the RRCV meets the minimal criteria. The view is stakes- and tradeoff-sensitive, building stakes and tradeoffs directly into the injunction to promote value. This allows the view to make a range of novel and plausible normative predictions, such as explaining 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 the RRCV.

The next chapter develops a second argument for the reason-responsive con-sequentialist view: the explanatory argument. The explanatory argument defends the RRCV based on its ability to deliver precise, well-motivated and unifying explanations of a variety of normative phenomena. Together with the argument from minimal criteria, the explanatory argument will complete my positive case for the RRCV.

6

The explanatory argument

6.1 Introduction

The bedrock of global consequentialism is its consequentializing program. Consequentializing aims to construct precise, plausible, unifying, and non-arbitrary consequentialist explanations for normative judgments and principles, focusing especially on claims that have eluded other theories or which seem to resist consequentialist analysis. A successful consequentializing program constitutes a programmatic *explanatory argument* for consequentialism on the basis of its explanatory fruits. Making out that explanatory argument will be the business of this chapter.

We saw in Chapter 3 that consequentializing must be done with care. Some modern consequentializers use liberalized readings of consequentialist mainstays such as promotion, outcome and value to argue that all, or nearly all normative theories can be consequentialized. For example, modern consequentializers might say that it is wrong to cut up one person to save five, because the outcome of this action would be a killing and a killing is worse than five lettings-die (Setiya 2018). Many nonconsequentialists have thought that these programs do not ground a strong explanatory argument for consequentialism, for example because they may trivialize consequentialism (Betzler and Schroth 2019; Tenenbaum 2014) or abandon its most powerful explanatory claim: that the good is prior to the right (Schroeder 2007).[1]

I don't want to take a stand on the correctness of these criticisms. But my program will not be subject to them. To the best of my knowledge, the arguments in this chapter appeal only to intuitive notions of value, outcome, and promotion. I avoid more controversial developments in the theory of value, such as the appeal to agent-relative value (Schroeder 2007). I understand promotion in what I hope is the ordinary sense of bringing outcomes about, and take outcomes to be what is brought about by actions in this sense. While the matter is not for me to judge, I hope that readers will agree that the arguments in this chapter interpret consequentialist mainstays in a natural sense that can ground a nontrivial and illuminating explanatory program.

---

[1] Trivialization would scuttle the explanatory argument insofar as a trivial explanation is taken to provide little or no explanation at all. Losing the priority of the good to the right would be a problem if consequentialists aim to provide grounding explanations for normative phenomena, and take claims about the good to ground claims about the right.

To some extent, the explanatory argument spans this entire book. If Chapter 7 is correct in claiming that the reason-responsive consequentialist view can deliver plausible rationalizing explanations of many challenged areas of human cognition, then the success of these explanations provides some evidence for the RRCV. Likewise, if the claims about bounded rationality made in Chapter 2 are plausible, and if Chapters 7–8 show how the RRCV can vindicate them, then this adds to the explanatory evidence for the RRCV. Many of the motivations given in Chapter 3 for adopting components of the RRCV were also explanatory claims.[2]

My aim in this chapter is to advance the explanatory argument by showing how the reason-responsive consequentialist view provides precise, plausible, unifying and non-arbitrary explanations of three inquiry-related normative phenomena that have resisted explanation by other theories. In Section 6.2, I show how the RRCV grounds and generalizes norms of clutter avoidance, which forbid agents from inquiring into trivial matters. Section 6.3 shows how the RRCV redescribes and vindicates many recent claims linking friendship to rational belief while stripping these claims of their most controversial or revisionary implications. Section 6.4 argues that the RRCV gives detailed, precise and plausible verdicts about the extent of logical omniscience required of bounded inquirers. Section 6.5 concludes.

6.2 Clutter avoidance

If you are like most people, you believe the following: the sky is blue. From this belief, you can deduce many others. For example, either the sky is blue or today is Quine's birthday. Must you spend your days drawing out each trivial logical consequence of your beliefs?

Many philosophers have followed Gilbert Harman in thinking that you should not do this. To clutter your mind with trivialities would be an irrational waste of time. Harman captured this thought in a principle of Clutter Avoidance.

Clutter Avoidance (CA) One should not clutter one's mind with trivialities. (Harman 1986, p. 12)

Harman gave three arguments for CA. First, *wasted effort* is expended in deducing trivialities. Second, due to our *limited retrieval capacity*, cluttering long-term memory with irrelevant beliefs makes it more difficult to retrieve relevant

---

[2] For example, we saw that reason-responsiveness allows us to explain why it is licit to pass between oughts and rational requirements in normative theorizing, and that information-sensitivity can be invoked to explain a range of normative and semantic phenomena.

beliefs from memory. Finally, our memories have *limited storage capacities* for information.

While the argument from limited storage capacities has met with mixed reactions, most philosophers have thought that there is something plausible about CA and that considerations of wasted effort and limited retrieval capacity are an integral part of the story.[3] Philosophers have disagreed about the contents of CA: what precisely does CA require? Philosophers have also disagreed about the explanation for the truth of CA. In this section, I review two prominent views about the contents of CA and its grounds, then develop an alternative consequentialist view.

Jane Friedman (2018) takes Clutter Avoidance to concern *junk beliefs*.[4] Friedman begins with the notion of a junk subject matter. Subject matters are junk for agents whose interests would not be served by knowing about them. More precisely, subject matter x is junk for agent S in world w at time t just in case S has no interest or desire served by having a belief about some proposition contained in x. Junk propositions are members of junk subject matters. Friedman takes Clutter Avoidance to be a prohibition against forming junk beliefs.

(**CA$_F$**) Necessarily, if p is junk for S at w, t, then S ought not believe p at w, t.

Friedman accepts Harman's motivations for CA$_F$. She also takes on board Harman's view that CA$_F$ is not a first-order normative claim, but rather a meta-normative constraint on acceptable norms of belief revision. By this, Friedman means that norms of belief revision that conflict with CA$_F$ cannot be genuine norms.

Note that *CA$_F$*, like *CA*, is phrased as a claim about oughts rather than rational requirements. The claim that rationality is deontic explains why this is not a conceptual mistake: we can pass between the coextensive categories of what agents ought to do and what they are rationally required to do. For this reason, we can meet Friedman halfway and speak in this section about how agents ought to inquire without risk of equivocation.

While I will take on board substantial portions of Friedman's view, the view has three limitations that a consequentialist approach helps resolve. First, inquirers should not merely be concerned with their own interests but also with the interests of others. Suppose that Sharky Sue is a CEO who is genuinely unconcerned with the welfare of her workers and benefits financially from exploiting them. Should Sharky Sue inquire into her workers' welfare? According to CA$_F$, Sue should

---

[3] The trouble with the argument from storage capacities is that many memory researchers think we are in limited danger of running out of memory storage (Michaelian 2011).

[4] There is an instructive parallel between Friedman's instrumentalist approach and Stich's (1990) pragmatic approach, which we met in Chapter 5. That discussion illustrates some other advantages of consequentialism over instrumentalism.

not. The welfare of Sue's workers is a junk subject matter for Sue since it would serve none of Sue's interests to have beliefs about her workers' welfare. But Sue's unconcern for her workers' welfare does not absolve her from the obligation to learn about their welfare any more than it absolves Sue from the obligation to improve her workers' welfare. We have no more normative license to be selfish in inquiry than in any other activity.

Second, although wasted effort and limited retrieval capacity are compelling motivations for CA_F, we need to explain why these considerations are normatively relevant to inquiry. As Friedman notes, many 'epistemic purists' take such considerations to be irrelevant to rational inquiry. A compelling response will not merely deny the epistemic purist's contentions, but also substitute an alternative picture of rational inquiry on which these considerations are normatively relevant.[5] We will see at the end of this section how consequentialism captures the normative relevance of wasted effort and limited retrieval capacities. We will also see at the end of Section 6.4 how consequentialism captures three further normative considerations driving theorists like Harman to adopt Clutter Avoidance.

Third, we need to clarify the objects to which CA_F applies. Although CA_F is stated as a principle governing belief states, Friedman clarifies that CA_F governs belief revision—that is, the process of inquiry by which our beliefs are revised. Clutter Avoidance has been applied to at least three separate processes of belief-revision: initial investigation, encoding of beliefs in long-term memory and forgetting. By separately specifying the demands of Clutter Avoidance on each of these processes we can clarify its contents and motivations. In each case, I will argue that in addition to any meta-normative content, CA_F generates specific first-order normative requirements on investigation, encoding and forgetting.[6] In formulating these norms I will leave open the conditions under which a belief counts as junk, since on my own view the junkiness of a belief is not essentially tied to a believer's interests.[7]

Suppose you are walking down the street and you notice that p: there is a ladybug on your neighbor's rosebush. As a result, you form the junk belief that p. Have you done something wrong? Plausibly not. We form junk beliefs of this sort all the

---

[5] Friedman (2020) provides one such picture: all epistemic norms are zetetic norms, and all zetetic norms are epistemic norms. We saw in Chapter 5 that the RRCV makes room for this view, but also opens the possibility of weaker views which allow a substantive distinction to be drawn between epistemic and non-epistemic norms.

[6] Friedman's argument against a first-order construal of Clutter Avoidance draws on her view that Clutter Avoidance is interest-sensitive, which I deny. Friedman is also concerned that first-order Clutter Avoidance norms will conflict with existing epistemic norms, but this is not clearly the case if existing norms are taken as constraints on belief states rather than processes of inquiry, given a sharp level separation between belief and inquiry.

[7] For consequentialists, 'junk' is an evaluative term in the standard axiological sense of the word. Junk beliefs are beliefs without significant value or disvalue. In this usage, 'junk' is opposed to 'significant,' and the junkiness or significance of a belief is a matter of degree. The turn to a graded notion of junkiness is needed to block trivialization, since few beliefs are utterly without value.

time, and it would be more effort than it is worth to prevent ourselves from forming them. But it would have been wrong for you to search the rosebush for ladybugs unless you were curious or wanted to protect the roses. This suggests that there is a clutter-avoidance norm on investigation driven by considerations of wasted effort.

> **Junk Non-Investigation (JNI)** Agents should not expend effort investigating junk subject matters.

You have not violated JNI because you did nothing more than notice that p in passing. It would have cost you more effort to suppress the junk belief than to form it.

Now that you have formed the belief that p, should you encode that belief in long-term memory? Some philosophers think that memory encoding is involuntary and hence inapt for rational assessment. However, at the very least it should be conceded that there are norm-governed voluntary processes that influence the depth and likelihood of encoding beliefs in long-term memory. For example, you can increase your probability of recalling a person's name by verbally repeating it after meeting them, and decrease the probability of successful encoding by distracting yourself. In the case of junk beliefs p, it is plausible at least that you should take cognitively inexpensive steps to ensure that p is not encoded, for example not staring at the ladybug. Otherwise p will clutter later memory searches, making it more difficult for you to make effective judgments and decisions based on information stored in memory. This motivates another first-order Clutter Avoidance norm.

> **Junk Non-Encoding (JNE)** Agents should take steps to ensure that junk beliefs are not encoded in long-term memory.

Plausibly, you have not violated JNE if you merely notice the ladybug in passing and end up encoding the belief that p in long-term memory.

But now that p is encoded in long-term memory, all is not lost. We can and do preferentially forget junk beliefs in order to declutter memory retrieval. Here is how the trick is done (Anderson and Schooler 1991). Junk beliefs are unlikely to have been accessed frequently or recently since they are irrelevant to most memory queries. By preferentially forgetting memories according to recall frequency and recency, we can rid ourselves of junk beliefs over time. Again, many philosophers think that memory is inapt for normative assessment, but this opinion is increasingly controversial: we regularly criticize agents for failures to remember (Carr 2015; Smith 2005) and a growing number of philosophers and psychologists develop rationality requirements on memory (Bernecker 2018; Schooler and Hertwig 2005).

I want to side-step the question of which steps toward forgetting are sufficiently volitional to be apt for rational assessment. Most philosophers accept that some forgetting-related steps can be the subject of rational requirements. For example, we can be rationally required not to intentionally mentally replay an unfortunate episode that we would like to forget because replaying the episode decreases the likelihood that we will forget it. Hence without taking a stand on the extent of rational requirements on forgetting, we can see Clutter Avoidance as imposing the following norm:

> **Junk Forgetting (JF)** Agents should take steps to ensure that junk beliefs are forgotten.

If you are like most people, you will soon forget all about the ladybug on the rosebush. That is a good sign that you have complied with JF.

Summing up, we need to develop an account of Clutter Avoidance that is unselfish, explains why standard motivations for Clutter Avoidance are good ones, and separately accounts for the first-order normative requirements imposed by Clutter Avoidance. A good start at this project is made by Kourken Michaelian (2011). Michaelian argues that the function of memory is to provide a manageable quantity of currently relevant information in a timely manner. These considerations could be taken to directly support JF and JNE, although JNI would need to be separately motivated by appeal to wasted effort. But this is not what Michaelian does. Michaelian argues that Clutter Avoidance can be supported by a purely intellectual norm that ranks belief states by the ratio of junk to non-junk beliefs contained in them. Roughly, the norm is:

> **Ratio (RT)** Let B, B' be belief states which differ only in that B' contains some junk beliefs that B does not contain. Then B' is epistemically inferior to B.

Michaelian motivates RT by arguing that we care intrinsically about the ratio of junk to non-junk beliefs contained in a belief state. I must confess that I find it opaque why agents should care about such a thing. It is perfectly intelligible why agents would be averse to spending effort moving from B to B', or why they would be worried that B' clutters later memory retrieval. It is also intelligible why agents would think that B' is not much epistemically better than B since the junk beliefs in B' are not very important. But why should the ratio of junk to non-junk beliefs contained in an agent's belief state count intrinsically against it? RT is an ad hoc principle invented to deliver Clutter Avoidance norms that are best explained by the very arguments that Harman took to motivate them.

A consequentialist approach to Clutter Avoidance fares better. Consequentialism explains the normative importance of limited retrieval capacities. As

Michaelian correctly notes, in memory retrieval agents aim to retrieve manageable quantities of relevant information in a timely manner. We do this in order to make good decisions and form correct judgments on the basis of information retrieved, because on a rich axiology it is better to make good decisions and judgments than bad ones. The need to declutter retrieval motivates JNE and JF.

Considerations of wasted effort are also normatively relevant to our initial investigations. It is wrong to waste effort investigating junk subject matters, even if the resulting beliefs will not be encoded in long-term memory. Wasting effort is wrong because that effort could be better spent on other pursuits. Limited retrieval capacities are also relevant insofar as junk investigations may, against our wishes, lead junk beliefs to be encoded in long-term memory, cluttering retrieval. Together, wasted effort and limited retrieval capacities provide good support for JNI.

In this way, consequentialism explains the normative relevance of Harman and Michaelian's original motivations for Clutter Avoidance norms. Consequentialism grounds the correct first-order normative consequences of Clutter Avoidance and does so without introducing unmotivated epistemic principles. Further, consequentialism is unselfish. Sharky Sue ought to investigate and take steps to remember facts about her workers' welfare because knowing these facts puts Sue in a position to improve the lives of her workers.

We have seen that consequentialism explains the nature and grounds of Clutter Avoidance norms, capturing the most plausible claims made by existing views while avoiding their defects. In the next section, we will see how consequentialism accounts for a second class of cases: duties of friendship.

6.3 Duties of friendship

Humans are social creatures. Because we are social creatures, friendship plays a central role in flourishing human lives. Our friendships generate normative obligations, for example, to care about and promote the welfare of our friends. Many philosophers take as a minimal constraint on the acceptability of normative theories that they should be compatible with the demands of friendship.

Sarah Stroud (2006) and Simon Keller (2004) have argued that friendship imposes normative demands on belief. For example, Keller recounts the following incident from 'Friends':

> **(Friends)** Joey gets a new job. Joey asks his friend Chandler if this job will be the big break he desires in his career. Chandler says that it will not. Chandler's disbelief leads to a falling-out between Joey and Chandler. Joey is upset not only that Chandler would voice such an opinion, but also that Chandler would hold it. Joey feels that his friend Chandler ought to think better of Joey's job prospects because Chandler is Joey's friend.

If Joey is correct, then what Chandler ought to believe is influenced by the fact that Chandler is Joey's friend. Stroud and Keller argue that this and similar norms relating friendship and belief are incompatible with traditional epistemic norms.

These claims have not been universally accepted (Kawall 2013). Perhaps Chandler is only required to keep his opinions to himself or to suspend judgment rather than forming the unjustified belief that Joey will succeed. And it is not obvious that the epistemic partiality due to friends is incompatible with traditional epistemic norms. For example, Stroud considers a man named Sam who is reputed to have slept with a woman and then refused to return her calls. Stroud argues that Sam's friends should expend increased effort gathering and scrutinizing evidence about Sam's behavior before believing that Sam has acted reprehensibly. But there is no clear epistemic norm against devoting additional scrutiny to evidence about Sam because he is your friend, any more than there is a norm against devoting additional scrutiny to the authenticity of a watch because you like watches or are making an expensive purchase.

What has been less discussed is that many of the normative claims made by Stroud and Keller concern rational inquiry rather than rational belief. Stroud's discussion suggests that Sam's friends are subject to two inquiry-related duties: to gather additional evidence and to keep inquiry open in the meantime. Similarly, Keller's central discussion concerns a woman Rebecca who is scheduled to give a poetry reading. Stroud argues that Rebecca's friends in the audience should listen with attention to the strengths of the poem; should devote less attention to weaknesses of the poem; and should seek interpretations that cast the poem in the best possible light. These duties concern the allocation of attention and the structure of reasoning, not belief. It is difficult to deny that friendship generates inquiry-related duties on evidence gathering, inquiry termination, attention, and reasoning. A prominent critic of Stroud and Keller grants all of these claims (Kawall 2013).

Friendship also generates many other inquiry-related duties. For example, we ought to inquire about the welfare of our friends and take steps to remember important details about their lives. These duties do not conflict with traditional epistemic norms since most traditional epistemic norms govern belief rather than inquiry.[8] Hence there is less pressure to deny them. But for the same reason, traditional epistemic theories such as evidentialism and coherentism will not explain the duties of friends during inquiry. How can these duties be explained?

The reason-responsive consequentialist view provides a natural and explanatorily powerful story. First, let us see how the RRCV accounts for the value of friendship. On an Aristotelian note, a life of shared activity and esteem among friends

---

[8] Process reliabilism may be an exception, insofar as it ties the rational status of a belief to the reliability of the process that produced it. But even here, process reliabilists could accept that friendship generates inquiry-related duties and hold that rational inquiry sometimes produces irrational beliefs. Consequentialists are not the only theorists who can benefit from a sharp level separation.

may be partly constitutive of well-being for humans as social animals. Friendship also has instrumental value of many kinds. Friendship is life-enhancing, giving friends purpose, energy and belonging. Friendship is an excellent tool for moral improvement, being based on esteem for the virtues of friends. And friends share information, undertake shared endeavors and care for one another.

On a consequentialist view, this axiological story extends to account for the inquiry-related duties of friends. For example, Rebecca's friends should attend to the poem's strengths so that they can support Rebecca with sincere praise after a daunting public reading. Focusing on the poem's strengths will also strengthen their esteem for Rebecca, preserving a friendship based on esteem for virtue. Sam's friends should gather evidence and prolong their inquiries because concluding that Sam has acted badly will rupture their friendships, and it is worth making reasonable efforts to ensure that friendships are not broken off unnecessarily. Similar arguments can be made to recover other inquiry-related norms.

Sometimes it is thought that consequentialism imposes normative demands which are incompatible with friendship. For example, friends should aim directly to promote their friends' welfare, not merely aim to promote the impartial good. If friends always made decisions by considering the impartial good rather than the immediate needs of their friends, they would be cold, distant and alienated from their friends (Stocker 1976; Williams 1973).[9] But consequentialism does not require agents to aim at promoting the impartial good (Parfit 1984; Railton 1984). If important goods such as friendship can only be realized by allowing ourselves to be directly moved by the needs of our friends, then that is what consequentialism instructs us to do.

Perhaps the problem is that consequentialism advises us to 'trade up,' leaving friendships behind in order to obtain more or better friendships (Badhwar 1991). But it is not so clear that consequentialism has this consequence. Betraying friends causes deeply personal, cutting forms of harm that are difficult to outweigh (Norcross 1997). Older friendships may also tend to be more fulfilling and valuable than new ones. But at the same time, we need to make room for the fact that agents should sometimes trade weak friendships for strong ones, or few friendships for many. Most of us have taken actions that foreseeably cost us some friends in order to gain others, and a natural justification for these choices is that the gained friendships were more valuable than those we lost.

This last remark suggests another virtue of the consequentialist account: it does not generate implausibly strong duties toward friends. Consider again the case of

---

[9] These cold, calculating agents would also take a long time to make up their minds and would be moved largely based on ill-informed estimates of impartial value rather than well-informed estimates of the needs of their friends.

Sam, who is reputed to have slept with a woman and then refused to return her calls. Stroud proposes that:

> What other people might classify as compulsive womanizing on Sam's part, [Sam's friends] might see as irrepressible but fickle enthusiasm and appetite for female charm in all its many varieties.   (Stroud 2006, p. 508)

Where others may call Sam a rascal, Sam's friends might say:

> There's never any artifice with Sam. You know where you stand with him: if he doesn't want to see you, he makes that clear. There's no false politeness, no pussyfooting, no hypocrisy, no stringing you along—Sam's too genuine for any of that.   (Stroud 2006, p. 508)

I think that we should be careful in drawing such strong conclusions.

Sam's 'fickle enthusiasm' and lack of 'pussyfooting' may represent a pattern of reprehensible and abusive behavior toward women with severe consequences. To valorize Sam's behavior in this way normalizes and protects behavior that may constitute serious abuse. Sam's friends are in an especially privileged position to identify and correct Sam's behavior, or to report Sam's behavior if Sam refuses to change. All of this tells in favor of relatively unbiased inquiry by Sam's friends. Similar conclusions apply across a range of cases. For example, if Sam's friends are hesitant to recognize when Sam is addicted to narcotics, then they will be unable to provide Sam with the help that only a friend can. Strong biases toward friends in inquiry constitute a form of epistemic protectionism that often benefits no one.

Finally, the RRCV can be used to protect rather than deny traditional non-consequentialist accounts of rational belief. By adopting a strict level-separation (Chapter 3) between norms of belief and inquiry, the consequentialist can accept friendship-related duties to inquire while denying, if she wishes, that friendship has any bearing on rational belief. In this way, my consequentialist view may be an excellent way to preserve rather than overturn traditional epistemological intuitions.

In this section, we have seen that the reason-responsive consequentialist view provides a flexible and plausible account of inquiry-related duties toward friends. Consequentialism does not require friends to be alienated from their affections or to trade friends like playing cards. Consequentialism avoids generating implausibly strong protectionist duties toward friends, and can be used to protect traditional epistemological views from the charge that they neglect friendship. In the next section, I argue that the RRCV can also be used to characterize requirements of logical non-omniscience.

6.4 Logical omniscience

How does logic constrain rational inquiry? On some views, rational inquirers are subject to norms of logical omniscience such as the following:[10]

(Deductive Closure) If a rational agent believes that $p_1, \ldots p_n$, and $p_1, \ldots, p_n$ entail q, then she forms the belief that q or abandons her belief in one of the p_i's.

(Consistency) If a rational agent believes an inconsistent collection of claims $p_1, \ldots, p_n$, then she abandons her belief in one of the p_i's.

Authors studying rational inquiry are nearly unanimous in their opposition to norms of logical omniscience (Harman 1986; Hintikka 2001).

At least three reasons are given for rejecting norms such as Deductive Closure and Consistency. First, full compliance is *beyond our abilities*. We cannot, for example, deduce all logical consequences of Peano Arithmetic. Second, it is often *too demanding* to comply with omniscience norms even when we could do so. To infer from p to $p \vee q$ is within our abilities, but often a waste of time.[11] In fact, Harman took such inferences as his original motivation for Clutter Avoidance. Finally, as Harman (1986) notes, the laws of logic are *principles of implication*. They tell us what follows from what. Principles of implication are not *principles of revision*, telling us how to revise our beliefs. We can propose principles of revision that are motivated by the laws of logic, but the status of these logical laws as principles of implication is no reason to accept them as principles of belief revision. Indeed, it would be surprising if there were a straightforward correspondence between the very different questions of what follows from what, and which processes rational agents should use to revise their beliefs.

It might be thought that weakened forms of Deductive Closure and Consistency survive these criticisms. For example, we could restrict both principles by requiring that the entailments in question be foreseeable or beneath a certain logical complexity. But such weakenings are at once too demanding and not demanding enough. These weakenings are too demanding because simple and

---

[10] MacFarlane (manuscript) and Steinberger (2019) term these 'bridge principles.' While a full discussion of their views is beyond the scope of this book, the move toward attitudinal restriction in Steinberger (2019) will be covered later in this section. Some other important issues, such as the distinction between narrow- and wide-scope readings, go beyond the focus of this book.

[11] What happens to this example on views which hold that believing p necessary entails believing $p \vee q$? I think that such accounts are probably not best suited to accounting for bounded rationality, because they make it hard to understand phenomena such as reasoning from p to $p \vee q$. However, readers who think that belief is not (of necessity) fully closed under logical implication are welcome to substitute their favorite trivial logical inference. And readers who think that belief is necessarily closed under logical implication are welcome to substitute their favorite example of a good but trivial non-logical inference instead, although this would require slightly rewriting the upshot of this section. Thanks to a reviewer for pressing me to address these questions.

foreseeable inferences such as the inference from p to $p \vee q$ may do nothing but waste effort and contribute to mental clutter. And they are not demanding enough because they make no room for stakes.[12] For example, humans make a variety of logical errors in evaluating important financial decisions, and these errors are enormously costly (Benartzi and Thaler 2007; Johnson et al. 1993). We could reduce the frequency of such errors by employing more demanding forms of reasoning which draw less foreseeable and more complex inferences and detect or avoid additional inconsistencies. But nothing in the weakened forms of Deductive Closure or Consistency explains why we ought to do so.[13]

Some authors respond by saying too little about the logical requirements on rational inquiry. For example, Christopher Cherniak weakens Consistency to say that:

> If an agent has a particular belief-desire set, then if some (but not necessarily all) inconsistencies arose in his belief set, he would eliminate them.
>
> (Cherniak 1981, p. 172)

This is not false, but it does not tell us what we would like to know. We want to know which inconsistencies a rational inquirer will detect and eliminate. Similarly, Harman weakens Consistency to hold:

> One has a reason to avoid believing things one recognizes to be inconsistent.
>
> (Harman 1986, p. 18)[14]

Again, this is too weak. We want to know which inconsistencies a rational inquirer should recognize, not merely that she has a reason to rid herself of inconsistencies once they are recognized.[15]

---

[12] For discussion, see Chapter 5.

[13] It might be objected that the threshold for complexity or foreseeability is context-sensitive. But now we need an account of how and why these thresholds vary, and here consequentialism seems a likely candidate.

[14] For a similar principle see Steinberger (2019). Note that Steinberger takes this principle to be a *directive*, offering first-personal guidance, rather than an *evaluative* third-person standard. If that is right, then we need not disagree since consequentialism is an evaluative claim. However, Steinberger later claims that the demandingness of inquiry is not evaluatively relevant. Here we do disagree.

[15] Additionally, we might draw on recent work in the semantics and metaphysics of belief. This work has sought to characterize ways in which the beliefs held by bounded agents may be hyperintensional, a necessary component of any account on which agents are not logically omniscient. Work on the semantics and metaphysics of belief can inform normative theorizing in at least three ways. First, it tells us how normative theories must be formulated: while the textual discussion treats belief as propositional, a full account will probably replace propositions with some other object. Second, it constrains the shape of normative theory: if belief in p necessarily involves belief in q, then the right normative theory cannot permit belief in p without belief in q (Stalnaker 1984). Finally, work in metaphysics and semantics suggests constraints such as relevance (Anderson and Belnap 1975; Dunn and Restall 2002) and subject matters (Fine 2020; Yablo 2014) which may bear on the question of which inferences are rationally required. At the same time, the primary object of these discussions has been

At this point, we might despair of saying anything substantive and true about the logical requirements on rational inquiry. But that would be a mistake. To see the way forward, consider the game of chess. Most chess players evaluate moves at least partly by considering candidate continuations, sequences of future moves that could result from a given position. Logically omniscient agents would not need to generate and examine novel continuations, since the existence of each continuation follows by first-order logic from the agent's knowledge of the present position and the rules of chess. But for non-omniscient agents, generating new continuations is a way of increasing the deductive closure of their belief state. Here we can ask the natural normative question: when should agents consider additional continuations, and which move sequences should agents search for continuations of?

These questions can be most precisely posed for agents whose computational abilities and memory are fully specified. A great deal of work has been done on these questions, including work by bounded rationality pioneers (Chase and Simon 1973; Newell et al. 1956; Newell and Simon 1965; Simon and Gilmartin 1979).[16] Some answers on offer are explicitly consequentialist. For example, agents might engage in consequentialist meta-reasoning, assigning to each position that they are aware of an expected value of searching for continuations beginning at that position and taking the searches with highest expected value (Lieder and Griffiths 2017; Russell and Wefald 1991). The expected value of search tracks the expected improvement to the player's position, and perhaps also intellectual achievements such as gained understanding, less the expected cost of deliberation time. Most authors hold that explicitly calculating expectations is more costly than it is worth. These authors instead propose algorithms for pruning and extending trees of possible continuations.[17] Although these algorithms do not explicitly calculate expectations, most scholars have understood them to have an explicitly conse-quentialist justification. If agents ought to evaluate some continuations rather than others or halt search rather than continue, that is because doing so will have the

semantic and metaphysical rather than normative, and most authors would take these discussions to fall short of providing a full account of the rationally required degree of logical omniscience. Although it may be possible to press some accounts into normative theories—requiring, for example, full logical omniscience within but not across subject matters—these accounts may face similar worries to those mentioned above, such as demandingness and stakes-insensitivity.

[16] It is no accident that the founders of bounded rationality took an interest in chess. An excellent strategy for working out the consequences of bounded rationality norms is to begin with a precisely specified activity such as chess, then analyze the cognitive capacities of human or non-human players, as a result of which we can say in great detail exactly what bounded rationality norms require. If those implications are plausible, then that is some reason to think we are on the right track. By contrast, theories that have unclear or implausible implications in well-specified contexts may be regarded with some skepticism.

[17] An accessible and effective example is alpha-beta pruning in minimax decisionmaking (Knuth and Moore 1953). Some recent algorithms function quite differently (Silver et al. 2018).

best impact on the player's odds of winning the game or their understanding of chess play.

I take this discussion to reveal two things. First, there is often a great deal that can be said about logical requirements on rational inquiry. In the special context of evaluating possible continuations during chess play, there is a well-established body of research offering specific and detailed characterizations of the degree of deductive closure that agents should achieve given constraints imposed by their cognitive capacities and deliberation time. Second, at least in the special case at hand, norms of inquiry are explicitly consequentialist: agents should inquire in the ways that are expectedly best, which in this case amounts to doing whatever makes them most likely to win the game. Features such as foreseeability and complexity matter only insofar as they bear on the possibility and cost of examining some continuations rather than others. There are many foreseeable combinations and even single moves that agents may rationally fail to consider.

Consequentialists think that logic plays the same normative role in chess as it does in all other inquiries. We can ask various questions about rational principles of revision. When should inferences be made using logical principles such as modus ponens? When may agents use inference patterns which sometimes produce inconsistent beliefs, and how much potential inconsistency should they accept? Each of these questions is to be answered on consequentialist grounds by citing the expected improvements to an agent's belief state, the importance of those improvements and the costs of making them. By way of illustration, we will see in Chapter 7 how consequentialists approach questions about conditional reasoning, and some have extended this approach to other domains such as a heuristic account of rational syllogistic inference (Chater and Oaksford 1999a; Oaksford and Chater 2007).[18]

A virtue of the RRCV is that it helps us to assess the importance of logical failures in context, balancing the importance of logical requirements against other cognitive desiderata. Certainly logical omniscience is instrumentally valuable, in helping us to form true beliefs and pursue our practical goals. And consequentialists may even grant that coherence has some intrinsic value. But the demands of logic are not sacrosanct, and they can be weighed against other goals.

Consider again the representativeness heuristic discussed in Chapters 1 and 5. The representativeness heuristic instructs us to judge the likelihood that object x belongs to category M by asking how representative x is of M. We saw in Chapter 1 that representativeness can produce incoherent judgments, for example

---

[18] A complication is that the Oaksford/Chater account takes a probabilistic construal of what previous theorists had assumed to be deductive reasoning. Similar tools could be finessed to generate normative theories of logical reasoning: for example, we might explore a normative theory of conditional reasoning on which agents assess the conditional $p \supset q$ by gathering evidence to discriminate between the hypotheses that $p \wedge \neg q$ and that $\neg p \vee q$. Working out and motivating such an account might be an important task for future research.

the judgment that Linda is more likely to be a feminist bank teller than a bank teller. I argued in Chapter 1 that this price may be worth paying for the sake of other goals, such as accuracy and cognitive efficiency.

Chapter 5 reminded us of another fact, which is seldom mentioned in scholarly discussions of representativeness: applied to Linda, representativeness is a straightforward form of inference by stereotyping. This observation helps us to put the importance of logic in its proper context. We may well be willing to run a small risk of making incoherent judgments about Linda. But we should not be willing to run a comparable risk of making biased judgments about Linda. Due to the harms perpetuated by stereotypes about feminists and women, we should classify this application of representativeness as irrational, not because it risks incoherence, but rather because it risks bias.

We saw in Chapter 5 that few competing views of rational inquiry can say this. Many accounts of rational inquiry cannot account for the irrationality of stereotyping, and most of the rest are unspecific about how goals such as bias avoidance are to be balanced against the demands of logic. By contrast, we saw that the reason-responsive consequentialist view, like its pragmatist cousin (Stich 1990) has a detailed story about how bias-avoidance is to be balanced against goals such as logical coherence. And we saw several reasons to prefer the RRCV to pragmatism.

A final advantage of the RRCV is that it captures the primary grounds on which logical omniscience requirements have been challenged: demandingness, cognitive abilities, and the distinction between principles of implication and principles of belief revision. For the consequentialist, the demandingness of principles of belief revision bears directly on their costs, and hence on their rationality. Rational requirements cannot be beyond our abilities, since rationality is deontic and in deontic matters, ought implies can. And principles of belief revision are to be justified by the costs and benefits of employing them, which need not coincide with principles of implication between propositions.

In this section, we have seen how the reason-responsive consequentialist view accounts for principles of logical non-omniscience. While theories of rational inquiry have been almost unanimous in their opposition to logical omniscience, existing theories struggle to develop detailed and plausible accounts of the required degree of logical omniscience. We used the example of chess play to see that there should be detailed principles governing the extent of logical omniscience that rational inquiry will achieve, and that the RRCV gives a natural explanation of what those principles are. We saw how this story can be extended to weigh the demands of logic against other desiderata such as bias avoidance. And we saw how the RRCV captures the primary grounds on which logical omniscience has been challenged.

6.5 Conclusion

This chapter developed the explanatory argument for the reason-responsive consequentialist view. The explanatory argument shows how the RRCV can be used to construct precise, plausible, unifying and non-arbitrary consequentialist explanations for a range of normative judgments and principles. If these explanations are carried out using the ordinary senses of consequentialist mainstays such as value, consequence, and outcome, then the explanatory success of the RRCV can be taken as evidence for its correctness. In this chapter, I showed how the RRCV can be used to explain three classes of normative phenomena.

Section 6.2 discussed norms of clutter avoidance. We saw that the RRCV allows us to tease apart three principles that have been grouped together under the heading of clutter avoidance and that it can ground those principles using the same plausible explanations which Harman originally proposed for norms of clutter avoidance.

Section 6.3 discussed norms of friendship. We saw that the RRCV relocates norms of friendship from belief to inquiry. In doing so, the RRCV carves out and vindicates a plausible core part of recent discussions of friendship in epistemology. At the same time, we saw that the RRCV prevents norms of friendship from falsifying traditional epistemic norms governing belief and avoids positing implausibly strong protectionist duties toward friends.

Section 6.4 discussed norms of logical non-omniscience. We used the example of chess to motivate the claim that there are specific and plausible norms relating logic to inquiry and to suggest that the RRCV provides a plausible account of those norms. In particular, the example of representativeness-based inferences revealed that the RRCV improves on many competing views by making room for the requirements of logic to trade off against other desiderata such as bias avoidance. And as in the case of clutter avoidance, we saw that the RRCV can capture without change all of the traditional reasons why theories of rational inquiry have resisted logical omniscience requirements.

Together with the argument from minimal criteria (Chapter 6.5), the explanatory argument completes the case for the reason-responsive consequentialist view. Our next and final task is to reap what we have sown. Parts 2–3 developed and defended an account of rational inquiry in order to ground and clarify the claims about bounded rationality set out in Chapter 2. Part 4 will use the RRCV to do the work it was introduced for. Chapters 7–8 show how the RRCV can be used to ground and clarify the five claims about bounded rationality introduced in Chapter 2. Chapter 9 uses the RRCV to revisit the relationship between bounded rationality and the Standard Picture. Chapter 10 concludes by applying the RRCV to draw lessons for recent work in the epistemology of inquiry.

PART IV
APPLYING THE ACCOUNT

7

Vindicatory epistemology

7.1 Reaping what we have sown

Chapter 1 introduced the Standard Picture of rationality, together with descriptive evidence that humans do not always obey Standard Picture norms. I considered two possible reactions to these empirical findings. On the one hand, we can blame the agent, retaining the Standard Picture and taking these findings to illustrate irrationality. On the other hand, we can blame the theory, using these findings to guide construction of a new account of bounded rationality.

Chapter 2 sketched a five-point bounded rationality program. We saw that to complete this sketch, we would need a theory of rational inquiry in order to clarify and ground its claims. Parts 2–3 constructed and defended a reason-responsive consequentialist account of rational inquiry.

The last item of business is to put this theory to work and see if it can do what it was introduced to do. In this chapter, I take up one part of the five-part bounded rationality program: vindicatory epistemology. Chapter 8 tackles the remaining four parts of the program. In each case, my aim is to show how the account of rational inquiry constructed in Parts 2–3 can deliver specific, plausible, and well-grounded explanations of what these normative claims come to in cases of interest to philosophers and cognitive scientists. Chapters 9–10 will conclude by reexamining the relationship between bounded rationality and the Standard Picture, and by applying the RRCV to generalize a recent turn in the epistemology of inquiry as well as to suggest the need for an analogous turn within practical philosophy.

7.2 Vindicatory epistemology

The business of vindicatory epistemology is to make good on our promise to blame theories rather than agents for many descriptive violations of the Standard Picture. The scholarly tide has turned away from the reactionary pessimism of the 1970s and 1980s toward cautious optimism about human rationality. Scholars from a wide range of perspectives have put forth compelling and detailed studies of the ways in which many of the most famous cognitive biases can be produced

by rational inquiry.[1] I find these studies persuasive, and I hope that by the end of this chapter you will too. If that is right, then a good normative theory should help us to understand the vindicatory arguments that these studies are making and to see why the arguments are good ones. Many traditional epistemological theories will not do this. Mine will. In this sense, it may be appropriate to treat vindicatory epistemology as an extension of the explanatory argument for the reason-responsive consequentialist view as well as a first application of that account.

Here is the plan. Section 7.3 lays out four desiderata on a vindicatory program. Section 7.4 introduces three strategies that my vindicatory program will follow in order to meet the desiderata. Sections 7.5–7.6 apply that program to two cognitive phenomena which are in need of vindication: anchoring and adjustment and conditional reasoning. Section 7.7 concludes.

7.3 Four desiderata for a vindicatory program

What should a good vindicatory program aim to accomplish? First and foremost it should be *vindicatory*. It should give compelling rationalizing explanations for large swaths of seemingly irrational patterns of judgment and decisionmaking. But life is not all sun and roses. A compelling vindicatory program should tell us not only what is right about human cognition, but also what is wrong with it. This demand generates our next two desiderata.

Chapter 3 argued that the proper normative stance toward many agents involves a type of evaluative tension. For example, a baker who saves too much bread for her family out of love has the right motives but takes the wrong action as a result. Our stance toward the baker is neither univocally positive nor univocally negative. Rather, we take up a torn and conflicted stance of evaluative tension toward the baker.

We feel the same evaluative tension toward participants in many heuristics and biases tasks. Consider a shopper who uses recognition as a search heuristic: she only considers products that she recognizes. This is a good way to narrow her options at the grocery store, but sometimes recognition leads our shopper astray. She walks straight past the store-branded bottles of still water and buys an expensive case of name-brand water, despite having loudly proclaimed the night before that she cannot taste the difference between competing brands of bottled water. How should we feel about this shopper? Plausibly, torn and conflicted. She may have inquired rationally: recognition is a good way to narrow searches at the

---

[1] See Icard (2018); Lieder and Griffiths (2020) in the Bayesian tradition; (Epley and Gilovich 2004, 2006) in the heuristics and biases tradition; and (Marewski and Schooler 2011; Pachur et al. 2011) in the fast-and-frugal heuristics tradition.

grocery store. But she has acted and intended irrationally: available information strongly indicated that the store-brand water would be a better bet. A good vindicatory program should *capture evaluative tension* in cases such as this one, motivating a conflicted evaluative stance and not a purely positive evaluation.

A third desideratum is *anti-Panglossianism*. Many attempts to vindicate human rationality are *Panglossian* (Stanovich and West 2000).[2] They imply that all, or nearly all human judgment and decisionmaking is rational. One Panglossian program claims that as a conceptual matter, humans must be competent reasoners and inquirers: irrationality could only ever amount to scattered performance errors in applying an underlying rational competence (Cohen 1981). Other Panglossians hold that our intuitions about rational judgment and decisionmaking are produced by reflective equilibrium beginning with our behaviors as inputs, and through this process we could not arrive at the conclusion that a substantial amount of human judgment and decisionmaking is irrational (Macnamara 1986; Sober 1978). These early positions led Daniel Kahneman to quip, not entirely without justification, that bounded rationality theorists see only two types of errors: "pardonable errors by subjects and unpardonable ones by psychologists" in constructing and interpreting experiments (Kahneman 1981, p. 340).

My program will not be Panglossian. With all due apologies to the good doctor, we do not live in the best of all possible worlds. Human judgment and decisionmaking is not nearly as bad as some have claimed, but it is far from perfect. Sometimes the best response is to accept that our judgment and decisionmaking exhibits patterns of irrationality. We can then turn to *meliorative epistemology* (Bishop and Trout 2004), which aims to improve human judgment and decision-making by understanding and counteracting patterns of human irrationality.[3]

The first three desiderata have said what a vindicatory program should accomplish. It should vindicate human rationality while capturing evaluative tensions and irrationalities when these are present. A final desideratum constrains the content of a good vindicatory program. Cognitive scientists have spent the past half-century escaping from the grip of experimental results which seem to paint human

---

[2] To be clear, I do *not* mean to endorse all or even most applications of this epithet in the literature, and I certainly don't want to endorse the way that bounded rationality theorists have been painted as Panglossians, *as opposed to* meliorists interested in improving human cognition. Bounded rationality theorists have been interested in meliorative epistemology from the very start (Galesic et al. 2009; García-Retamero and Hoffrage 2013; Gigerenzer and Hoffrage 1995; Hoffrage et al. 2000; Sedlmeier and Gigerenzer 2001). But I certainly do want to distance myself from the primarily conceptual, philosophically driven vindicatory programs cited in this paragraph.

[3] Actually, that is not quite right. We can perfectly well aim to meliorate *rational* cognition by improving the cognitive conditions that agents face, making more accurate strategies accessible to them. For example, it may be rational for busy doctors to process statistical information in whatever format that information is presented to them, but important for authors to present that information to doctors in a way that facilitates base-rate usage. On my understanding, melioration proceeds by making cognition more likely to possess desirable features such as accuracy, efficiency and explainability. Promoting rationality is one of many ways to do this.

rationality in a strongly negative light. A long series of theoretical arguments, empirical results, and modeling exercises have led most theorists to conclude that most human judgment and decisionmaking is more rational than previously thought (Bishop and Trout 2004). I think that these are good arguments, and they should not be wasted. The fourth desideratum is then *continuity* with existing theoretical and empirical discussions. Whenever possible, a good vindicatory program should capture existing vindications of human rationality in the scientific literature without changing their contents. When this is not possible, a good vindicatory program should build substantially upon existing empirical and theoretical discussions.

Sections 7.4–7.6 develop a vindicatory program with these four features. It is vindicatory, rationalizing many seemingly irrational patterns of judgment and decisionmaking. It captures evaluative tension in many of these cases. It avoids the Panglossian claim that all, or nearly all human judgment and decisionmaking is rational. And it is continuous with existing scientific discussions.

7.4 Three vindicatory strategies

In the previous section, I outlined four desiderata for a good vindicatory program. In this section, I introduce three strategies that my vindicatory program will employ and explain how each strategy stacks up against the desiderata.

When possible, I will argue that observed behaviors reflect rational inquiry leading to irrational belief or intention.[4] This is for example what happened to our recognition-driven shopper. Rational processes for narrowing options led her to reject the option she had most reason to choose. In addition to being vindicatory, this strategy captures a clear type of evaluative tension: we simultaneously judge that the agent has done something rational and something irrational. This strategy will often be applied as a continuous extension of existing scientific arguments. And it will be anti-Panglossian: I will pay clear attention to the boundary conditions of this explanation, showing where and why it breaks down as a vindication of observed behaviors. For example, recognition-driven search may be an irrational form of inquiry at the drug store, where generic brands often present substantial cost savings with little drop in quality.[5]

When this strategy is not available, I will use two other strategies. These strategies meet only some of the desiderata on a vindicatory program. My second strategy is to claim that empirical observations reveal no bias of any kind. This explanation is vindicatory and draws on existing scientific discussions, but it does

---

[4] This strategy relies crucially on the notion of level separation introduced in Chapter 3. Thanks to an anonymous referee for pushing me to clarify this application of level separation.

[5] For general discussion of the ecological validity of recognition, see Pachur et al. (2012).

not capture any type of evaluative tension and tends toward the Panglossian. My third strategy is to claim that empirical observations reveal a kind of irrationality that cannot be vindicated. This strategy demonstrates my commitment to an anti-Panglossian view of human judgment and decisionmaking. However, it is not vindicatory and captures no evaluative tension, so I will try to make limited use of this strategy.

Going forward, my primary strategy will be to claim that observed behaviors represent rational inquiry leading to irrational belief and action. When this is not possible, I will sometimes deny that any bias has been observed, and at other times grant the bias and deny that vindication is possible. In the next two sections, I apply these strategies to two classes of seemingly irrational behaviors in the judgment and decisionmaking literature.

7.5 Anchoring and adjustment

7.5.1 The origins of anchoring and adjustment

The heuristics and biases program originally focused on three heuristics which were claimed to underly much of human judgment and decisionmaking (Tversky and Kahneman 1974). One of these heuristics is anchoring and adjustment, which we met in Chapter 5. Anchoring and adjustment instructs agents to estimate a quantity by beginning with an *anchor* value, then iteratively *adjusting* their estimate to incorporate new items of information.

In the paper which launched the heuristics and biases program, Tversky and Kahneman (1974) proposed that humans often make judgments using anchoring and adjustment and offered a number of speculative applications of this heuristic. In the best-known example, participants were asked to spin a wheel of fortune. They then made a *comparative judgment*: is the percentage of African nations in the United Nations more or less than the number on the wheel? Participants then made an *absolute judgment:* what percentage of African nations are in the United Nations? Tversky and Kahneman found a strong *anchoring effect*: absolute judgments were biased toward the anchor value on the wheel. They argued that participants had answered the absolute question by a process of anchoring and adjustment, beginning with the number on the wheel as an anchor and making incremental, but insufficient adjustments toward the true value. This echoes a more general finding: adjustments away from an anchor are typically insufficient.

Tversky and Kahneman offered three further examples of anchoring and adjustment. For example, humans sometimes overestimate the probabilities of conjunctive propositions and underestimate the probabilities of disjunctive propositions (Bar-Hillel 1973; Nilsson et al. 2013). Tversky and Kahneman proposed that probability estimates are anchored on the probabilities of simple propositions,

then adjusted upwards toward the disjunctive proposition or downwards toward the conjunctive proposition. Because adjustments are typically insufficient, probabilities of conjunctive propositions will be overestimated and probabilities of disjunctive propositions will be underestimated.

These remarks were quite speculative. At no point had Tversky and Kahneman obtained evidence for a mental process of anchoring and adjustment, and it is somewhat implausible to suppose that the judgments described above were made by an explicit process of anchoring and adjustment. In the ensuing years, a search was launched for evidence that judgments are sometimes made by a mental process of anchoring and adjustment (Johnson and Schkade 1989; Lopes 1982). For two decades, none was found. Process-tracing measures yielded clear evidence of an anchoring effect: some judgments begin by taking an anchor value of an input. But until recently, no evidence of adjustment was found, and it is now agreed that neither of the cases described above involves anything like an explicit process of anchoring and adjustment.

As evidence for a heuristic process of anchoring and adjustment failed to materialize, research shifted toward anchoring effects which occur without any process of mental adjustment. In Section 7.5.3, I argue that these effects are not strong evidence of irrational inquiry. But first, I review the small class of cases involving genuine anchoring and adjustment. I argue that the evidence shows our inquiries in these cases to be highly rational and that the best explanation for why that is so involves an explicitly consequentialist model of rational inquiry.

7.5.2   Anchoring and adjustment

About twenty years ago, evidence emerged that humans may employ an explicit process of anchoring and adjustment to make judgments in a small class of cases (Epley and Gilovich 2001, 2004). We have already met one such case. Asked to estimate the date when George Washington was first elected president, many participants begin by considering the date 1776 when the Revolutionary War began, then adjust upwards to account for subsequent events. Similarly, agents asked to estimate the boiling point of water on Mount Everest often begin by considering the anchor value 212°F, adjusting downwards for the effects of increasing altitude. These cases differ from more theatrical examples involving spinning wheels of fortune in at least three relevant ways.

First, the anchors are *self-generated*. That is, they are provided by agents themselves during the normal course of reasoning. The anchors are not provided by experimental interventions such as spinning wheels. It is now thought that genuine anchoring and adjustment nearly always arises in response to self-generated anchors (Epley and Gilovich 2001). Hence there is scant opportunity for genuine anchoring and adjustment to be manipulated by spinning wheels or other attempts to provide misleading anchors.

Second, the anchors are *informative*. They convey information that is highly relevant to the problem at hand. The date of the Revolutionary War provides excellent information about the date of Washington's election, and the boiling point of water at sea level does the same for the boiling point of water on Mount Everest. While it might seem hopelessly irrational to estimate the composition of UN member states by anchoring on the value of a spinning wheel, there is no obvious irrationality in trying to estimate the boiling point of water at higher elevations by beginning with known facts about its boiling point at lower elevations.

Third, our judgments in these cases are often *highly accurate*.[6] This means that at first blush, anchoring and adjustment appears to be an excellent way to carry out the relevant inquiries, combining high accuracy with remarkable frugality.

Now it is true that in aggregate, judgments made by anchoring and adjustment display an anchoring bias. The data from the previous paragraph show that typical judgments made by anchoring and adjustment, while reasonably accurate, are slightly skewed in the direction of the anchor. Is this anchoring bias evidence of irrationality?

We might take anchoring bias as evidence of irrational belief. Anchoring bias occurs when adjustments have been insufficient to counteract the initial effect of the anchor, suggesting that only a moderate number of adjustments have been made, and hence that most relevant information has not been considered. This could be evidence that beliefs produced by anchoring and adjustment are often evidentially unsupported since they do not consider the entirety of an agent's evidence. But the same could be said about beliefs formed by almost any other process regularly employed by bounded agents. It is extremely uncommon for agents to consider all of their relevant evidence before making a judgment. It is for exactly this reason that theories of bounded rationality make the procedural turn: we ask not whether the beliefs that result from anchoring and adjustment are rational, but rather whether the process itself is used rationally.

Does the presence of anchoring bias show that a given instance of anchoring and adjustment was an instance of irrational inquiry? Not necessarily. There is broad agreement that the reason why adjustments are insufficient to counteract anchoring bias is that adjustment is effortful (Epley and Gilovich 2006; Lieder et al. 2018). Making enough adjustments to eliminate any trace of the initial anchor

---

[6] For example, Epley and Gilovich (2006) find that participants estimate Washington's election year to within 2.5 years, and the boiling point of water on Mount Everest to within 1.2°F. At the same time, it is important to stress that judgments in standard anchoring paradigms may also be reasonably accurate without any explicit process of mental adjustment. For example, Jacowitz and Kahneman (1995) find mean errors of 5% in estimating the height of Mount Everest, and 7% in estimating the average American's annual meat consumption. The most likely culprit for highly inaccurate estimates may well be lack of knowledge. For example, Jacowitz and Kahneman (1995) find mean estimates putting the number of female professors in Berkeley at just 83, compared to a true figure of 805 at the time, which may be primarily explained by the fact that respondents had limited information about the size and composition of university faculties. See Wilson et al. (1996) for evidence that knowledge mitigates anchoring effects.

would greatly increase the cognitive expense of anchoring and adjustment, often without significant increase in accuracy. To assess the rationality of inquiries made by anchoring and adjustment, we need to ask whether agents use anchoring and adjustment so as to strike the best balance between judgmental accuracy and cognitive effort during inquiry. Our best models suggest that they do.

We saw in Chapter 5 that the best-fitting model of anchoring and adjustment is an explicitly consequentialist model due to Falk Lieder and Thomas Griffiths (2018). This model represents agents as striking a near-optimal balance between accuracy and effort in determining how many adjustments to make. They do this by selecting a number of adjustments that minimizes, in expectation, the combined costs of cognitive effort and inaccurate judgment. Importantly, we saw that this model can be used not only to suggest that anchoring bias often results from rational anchoring and adjustment, but also to rationalize many other findings in the anchoring paradigm as well, such as the findings that monetary incentives often reduce anchoring bias (Simmons et al. 2010) and that anchoring bias increases as agents become busy with other tasks (Epley and Gilovich 2006).[7]

From this discussion, we learn two things. First, our best models suggest that observed anchoring bias in cases of genuine anchoring and adjustment is not evidence of irrational inquiry. Quite the contrary, we can explain not only the bare fact of anchoring bias but also many of its contours as the results of rational inquiry by bounded agents aiming to balance accuracy and effort in cognition.

Second, the best existing vindication of anchoring and adjustment is consequentialist. The Lieder and Griffiths model is an explicit application of expected utility theory. The effects of cognitive load and monetary incentives are predicted by letting both cognitive and noncognitive costs enter into expected utility calculations in the standard way. This is some evidence that a consequentialist theory is well-suited to the work of constructing a plausible and non-revisionary vindicatory epistemology.

7.5.3 Anchoring without adjustment

As it became increasingly clear that standard experimental paradigms did not involve a procedure of iterated anchoring and adjustment, research expanded to study a broader class of anchoring effects. What precisely is an anchoring effect? It is difficult to find a clear and specific definition of anchoring effects in

---

[7] And in fact, Lieder and Griffiths aim to rationalize *all* major regularities in anchoring and adjustment, including the effects of time pressure, alcohol consumption, anchor extremity, uncertainty, knowledge, and intrinsic accuracy motivation.

the literature, and as we will see this is no accident. The difficulty of offering a definition reflects the fact that several effects have been grouped under the same heading.

For most authors, an *anchor* is a value provided before making a judgment on an absolute scale and an *anchoring effect* occurs when that judgment is biased in the direction of the anchor. Here is a sampling of statements consistent with this view of anchoring effects.

> An anchor is an arbitrary value that the subject is caused to consider before making a numerical estimate. An anchoring effect is demonstrated by showing that the estimates of groups shown different anchors tend to remain close to those anchors.   (Jacowitz and Kahneman 1995, p. 1161)

> The anchoring effect is the disproportionate influence on decision makers to make judgments that are biased toward an initially presented value.
> (Furnham and Chu Boo 2011, p. 35)

> Anchoring is apparent in the assimilation of a numeric estimate to a previously provided standard.   (Mussweiler and Strack 1999, p. 139)

Although many authors require anchors or absolute judgments to be numerical, similar effects can be found in non-numerical contexts as well (Northcraft and Neale 1987).

This definition of anchoring effects is very broad since it is compatible with any of a number of distinct psychological phenomena. Indeed, an increasingly popular view is that there is no one anchoring effect, but several (Turner and Schley 2016). This is an occupational hazard of the heuristics and biases approach. The approach aims to identify biases so specific that they are almost always produced by a single heuristic strategy. Often this is a good approach, but sometimes it backfires. When it turns out that the specific bias is not typically caused by the intended heuristic, in this case anchoring and adjustment, things get messy. Biases such as anchoring come to be understood in a very broad sense that is compatible with a number of different cognitive processes, and great care is needed not to run distinct cognitive processes together. This much is conceded by Kahneman himself:

> The terms *anchor* and *anchoring effect* have been used in the psychological literature to cover a bewildering array of diverse experimental manipulations and results . . . The proliferation of meanings is a serious hindrance to theoretical progress.   (Jacowitz and Kahneman 1995, p. 1161)

If this is right, then the best we can do is to set out and study the various psychological processes that lead to anchoring effects and ask separately after the rationality of each one.

Many authors take it as obvious that anchoring biases are irrational biases if the anchor information is irrelevant to the judgment at hand. For example, Gretchen Chapman and Eric Johnson assert that "anchoring is clearly a bias when the starting point is random and uninformative" (Chapman and Johnson 1994, p. 223). By contrast, I think that careful attention to the mechanisms driving anchoring effects will vindicate most anchoring effects as the results of rational inquiry. I do not deny that anchoring effects typically result in irrational beliefs, insofar as they lead to evidentially unsupported beliefs with no clear non-evidential considerations to vindicate them. In this subsection, I review three mechanisms responsible for anchoring effects and argue that most are instances of irrational belief resulting from rational inquiry.

Response biases

Some anchoring effects do not change agents' beliefs at all. Rather, they distort the response scale on which agents report their beliefs (Frederick and Mochon 2012). If I show you a wolf, then ask you to point out an animal which weighs approximately 1,000 pounds, you will show me an animal slightly heavier than 1,000 pounds in order to emphasize that the animal you have picked is heavier than a wolf. But if I began by showing you an earthworm, you would point to an animal weighing less than 1,000 pounds because that is more than sufficient to make the salient contrast.

Distortions of response scales are not judgmental biases at all, since they do not change an agent's beliefs. Nor are they clearly irrational. The distortions I have mentioned are naturally supported by Gricean considerations of relevance and helpfulness. In fact, I don't know of any theorists who have even alleged that this sort of response bias could be irrational. So we will have to look elsewhere for evidence of irrational anchoring effects.

Numerical priming

Some anchoring effects are simple numerical priming effects. These anchors temporarily raise the availability and salience of the anchor value across the board, with an influence not limited to the question at hand (Wilson et al. 1996). Here I think there is no irrational inquiry to be found. Priming effects are nothing new, nor are they particularly interesting. Of course agents cognize differently when they are thinking about large numbers, just as they cognize differently when they are angry, thinking morally, or seated in a luxury car. There is no rationality requirement to cognize in the same way across situations, and some reason to think that rationality requires exactly the opposite.

For example, agents become increasingly risk-seeking when angry (Lerner and Keltner 2001), but this may be a beneficial response to threatening situations. Once we understand this tendency it can be manipulated by priming. You can insult your opponent across the poker table to lure them into making a bad bet. But I don't

think this shows that risk-seeking is an irrational response to anger. We could just as easily exploit agents who are not prone to risk-seeking when angry by harming them and expecting that they will not engage in costly retaliation. Since rationality is ecological, to show that a cognitive disposition is irrational it is not enough to show that the disposition can be exploited. All dispositions can be exploited when you understand how they work. We need to show that the disposition is ill-suited to the problems that humans typically face.[8]

What does this mean for numerical priming in anchoring effects? In general, it is a good strategy to allow exposure to high numerical values in the environment to temporarily increase your numerical judgments on compatible scales. Agents who failed to do this would increase their likelihood of ignoring evidence since they would not take explicit account of recently observed numerical values unless they had an extraordinary capacity for attention and an enlarged working memory. Priming effects demonstrate what we already knew: environments can be manipulated to provide misleading numerical primes. But this is no worse than any other manipulation of an agent's evidential situation. The bare existence of numerical priming in anchoring effects shows no more than that a good strategy for inquiry can be led astray by providing agents with misleading evidence. Demonstrating irrationality would require detailed criticisms of the ecological rationality of numerical priming in anchoring effects, and to my knowledge this feat has never been attempted.

Selective accessibility

Some anchoring effects act directly on an agent's beliefs with an influence confined to the task at hand. The best established of these is a *selective accessibility effect*: anchor-consistent information is made more accessible in estimating the target value. A series of papers by Thomas Mussweiler and Fritz Strack established the existence of selective accessibility effects as well as their cause (Mussweiler and Strack 1997, 1999, 2000). In the standard anchoring paradigm, participants are asked to make a comparative judgment: is the target larger or smaller than the anchor value? In answering the comparative question, participants adopt the working hypothesis that the anchor is equal to the target value and mount a hypothesis-consistent memory search for information consistent with this hypothesis. This hypothesis-consistent search makes anchor-consistent information more likely to be retrieved from memory, and hence more available

---

[8] This discussion raises the important issue of how dispositions should be individuated. For example, is the disposition in question the disposition to seek-risks-when-angry or the finer-grained disposition to seek-numerical-risks-when-angry, or the still finer-grained disposition to seek-poker-related-numerical-risks-when-angry? This question is important, because the latter dispositions may be less beneficial dispositions for agents to have. For now, all I have to say is that the individuation of dispositions is an important question that deserves a good deal of further study, and that in the meantime I will try not to individuate dispositions in suspicious or unnatural ways. Thanks to an anonymous referee and also to Jennifer Carr for pressing me on this.

when subjects are subsequently asked an absolute question: what is the target value? As a result, answers to the absolute question are biased toward the anchor.

Selective accessibility effects are the most promising candidates for irrational anchoring effects. But the case for the irrationality of selective accessibility effects is not strong. There are three ways in which it could be argued that selective accessibility effects are irrational. The first is to argue that memory search should not be hypothesis-driven. The second is to argue that memory search should be driven by a different hypothesis. And the third is to argue that memory search should seek hypothesis-inconsistent information rather than hypothesis-consistent information. The first road has not been taken, and for good reason: undirected memory searches are unlikely to turn up large quantities of useful information.

The second route, arguing that memory search should be driven by a different hypothesis, would be quite unfair if meant as a criticism of participants in anchoring studies. The experimental design provides a comparative question with the explicit intention of raising the salience of a given hypothesis: that the target value is equal to the anchor. So we do not learn anything about the mechanisms by which hypotheses are raised to salience from existing anchoring studies, and hence we are not in a good position to argue that participants tend to guide memory search using poor hypotheses. A few experimental paradigms find anchoring effects without asking comparative questions, but these studies tend to use informative anchors and hence there is no evidence for irrationality in these studies (Northcraft and Neale 1987). To establish irrational hypothesis selection, we would need evidence bearing on the ecological validity of the mechanisms by which information in the environment influences hypothesis selection. If irrelevant information frequently influences hypothesis selection for the worse, then we will have evidence for irrationality. But we cannot establish anything about the rationality of hypothesis selection by manipulating hypothesis selection toward a misleading hypothesis under artificial conditions.

The third route, arguing against hypothesis-consistent search, seems more promising. Isn't hypothesis-consistent memory search a type of irrational confirmation bias? But Mussweiler and Strack emphatically reject this strategy, at least insofar as the comparative question is concerned. They argue that confirmatory testing of the hypothesis that the anchor and target value coincide is an excellent strategy for settling the comparative question of whether the anchor is higher or lower than the target value. Roughly, their argument is the following. Agents seek diagnostic evidence bearing on the comparative question. A good way to get this evidence is to test the hypothesis that the anchor and target value coincide. Agents test this focal hypothesis with the goal of falsifying it, since the direction in which the most falsifying evidence has accumulated will provide a good answer to the comparative question.

Here Mussweiler and Strack invoke a surprising finding that motivated much of the vindicatory research on hypothesis testing summarized in the next section. Across a range of environments, searching for hypothesis-consistent information is a more efficient way to falsify a hypothesis than searching for hypothesis-inconsistent information (Klayman and Ha 1987). And that finding arose precisely in the context of confirmation bias, in an argument that many search strategies which look to be an irrational sort of confirmation bias are anything but. Hence Mussweiler and Strack conclude that observed performance on the comparative task is highly rational, and in particular that it is not a type of irrational confirmation bias.

Mussweiler and Strack note an unfortunate consequence of this result. The optimal hypothesis-consistent search strategy for answering the comparative question biases answers to the later absolute question because it makes anchor-consistent information more available in working memory. Does this constitute a type of irrationality?

Here I am not sure what to say. We are not often faced with comparative and absolute judgmental tasks in quick succession. We have learned that hypothesis-consistent testing is a good way of making comparative judgments, and it may well be that we apply this strategy to comparative judgments without reflecting on whether they will bias a later absolute judgment. It may also be that we do not learn to correct this error across repeated trials.

If this is right, then we may have discovered an instance of irrational inquiry under quite specific experimental conditions. But I will want to insist that this irrationality is not widespread, insofar as we rarely face such conditions. I will also want to suggest that this irrational inquiry could result from fully rational metacognition. Any form of metacognitive monitoring sufficiently detailed and responsive to detect and correct the mismatch between inquiry strategies and rarified situations such as this one would be far too cognitively costly for bounded agents to use.

Summing up, there are at least three types of anchoring effects. The first, response biases, are not effects on judgment at all. The second, priming effects, are not irrational in any clear way. The third, selective accessibility effects, exhibit at most a highly localized type of irrational inquiry which could nonetheless result from fully rational metacognition.

Putting this together with our previous discussion, we have seen that most behaviors originally attributed to anchoring and adjustment can be vindicated by the reason-responsive consequentialist view. These include the few genuine cases of anchoring and adjustment, together with three of the leading phenomena which have been grouped together as anchoring effects. Each vindication shows that what looks like irrational inquiry can be explained as an attempt to do something powerfully right, such as efficiently testing a hypothesis during memory search

or balancing accuracy and effort in cognition. Moreover, the vindication of these behaviors makes explicit and seemingly essential appeal to distinctive elements of a bounded consequentialist approach such as ecological rationality and expected utility maximization. That is a strong mark in favor of the view. In the next section, I show how a consequentialist view can be used to vindicate performance in a second area: logical reasoning.

7.6 Logical reasoning

7.6.1 Introduction

How good are humans at logical reasoning? Research on this question has concentrated on conditional reasoning. The bulk of attention has centered on two tasks.

In *conditional inference tasks*, participants are given a conditional statement and a single premise, then asked whether any conclusion follows. For example, the statements (if p, then $\neg q$) and q may be provided, in which case $\neg p$ follows by modus tollens. Performance on abstract conditional inference tasks is seemingly quite poor, with as many as half of participants affirming the consequent and denying the antecedent (Schroyens et al. 2001).

In the *Wason selection task*, participants are shown four cards, labeled respectively with an ace (A), king (K), two (2), and seven (7) (Wason 1966). They are told that each card has a number on one side and a face card on the other. Participants are asked which cards they would need to turn over to test the rule:

If a card has an ace on one side, then it has a two on the other side.

Wason intended this task to be a test of conditional reasoning. Labeling the cards as p (A), $\neg p$ (K), q (2), and $\neg q$ (7), respectively, the rule is that if p, then q. If this is read as a material conditional, it can be tested only by searching for falsifying instances in which p and $\neg q$ hold, hence participants should turn the ace and the seven. But fewer than 10% of participants make that selection (Oaksford and Chater 1994).

For several decades, these experiments were taken to show that humans are poor logical reasoners. This consensus has been questioned by a recent probabilistic turn in the psychology of reasoning (Elqayam and Over 2013; Oaksford and Chater 2020). Most human reasoning takes place under conditions of uncertainty. We rarely encounter exceptionless generalizations which can be expressed by the material conditional. We are usually interested in capturing probabilistic relationships and causal information, which are respectively captured by probabilistic and counterfactual readings of the conditional. On this understanding, tasks such as abstract conditional inference and Wason selection do not reveal poor logical

reasoning, because participants are not even attempting to do logical reasoning. These tasks rather reveal good probabilistic and causal reasoning which was wrongly assumed by experimenters to be bad logical reasoning.

In this section, I focus on probabilistic approaches to the Wason selection task. In Section 7.6.2, I outline an information-gain model on which behavior in the Wason selection task reflects rational evidence-gathering to discriminate between probabilistic dependency hypotheses. Section 7.6.3 extends the model to deontic versions of the selection task in order to bring out its consequentialist commitments. Although the information-gain model is popular and well-supported, it is controversial, and hence in Section 7.6.4 I review some evidence for the model. The bad news I save for Section 7.6.5: although the information-gain model shows behavior on the Wason selection task to be far more rational than previously supposed, it also highlights several clear irrationalities. I take this as an opportunity to make good on my anti-Panglossian commitments by owning up to patterns of irrationality in card selection.

7.6.2 The information-gain model

Mike Oaksford and Nick Chater have proposed an information-gain model of the Wason selection task (Oaksford and Chater 1994, 2007). On this model, participants select cards in order to gather evidence about the probabilistic relationship between p and q. There is a standard consequentialist account of duties to gather evidence on which evidence should be gathered in order to maximize the expected value of information received, less the cost of gathering it (Howard 1966; Stigler 1961). That is just what Oaksford and Chater propose that agents do.[9]

For simplicity, assume that agents are trying to decide between the *independence hypothesis* (M_I) that p and q are probabilistically independent ($P(q|p) = P(q)$), and the *dependency hypothesis* (M_D), that p and q are probabilistically dependent ($P(q|p) = 1$).[10] Because the problem is abstractly specified and divorced from action guidance, the value of information is its ability to reduce uncertainty about

[9] The information-gain model gains plausibility against the background of a broader probabilistic turn in the psychology of reasoning, which uses similar assumptions about conditional reasoning and other reasoning processes to generate a range of descriptively fruitful predictions that raise the possibility of taking human reasoning to be, if not perfect, then certainly more rational than previous logic-based accounts made human reasoning appear to be (Elqayam 2018; Oaksford and Chater 2007, 2020).

[10] I know of no place where Oaksford and Chater explicitly claim that this is a simplifying assumption, but I think they must intend it as such, and certainly on my normative view it had better be a simplifying assumption. We have reason to learn a great deal more about $P(q|p)$ than whether it exhibits full dependence or full independence, and indeed good reason to think that both hypotheses are false in most contexts. So for a consequentialist, we ought to maximize the informativeness of our overall view about $P(q|p)$, a fact which would greatly complicate the relevant mathematics and empirical modeling, but which must be treated as the basic normative fact to be captured.

the correct hypothesis and all uncertainty is equally important.[11] The expected value of each evidence-gathering action, namely turning a card, is the expected information gain—that is, the expected uncertainty reduction from turning the card. Consequentialists hold that agents ought to turn the cards with the highest expected information gain.

Oaksford and Chater's insight was that when p and q are both improbable, the cards with the highest expected information gain are p and q. Hence the information-gain model recommends that agents turn the p and q cards, and that is the modal response in abstract versions of the Wason selection task. Oaksford and Chater make the *rarity assumption* that agents treat p and q as improbable in the Wason selection task. They draw on earlier results suggesting that agents do, and should treat events as rare in most conditional and causal inference contexts (Anderson 1990). If this is right, then the information-gain model seems to vindicate the modal choice in Wason selection tasks. How much of observed behavior can it vindicate?

At this point matters become somewhat delicate. Oaksford and Chater's model relates choice to scaled information gains, which are expressed as a fraction of the possible expected information gains.[12] There is nothing irrational yet, but Oaksford and Chater did not find that participants always select the cards with highest (scaled or unscaled) expected information gain. They found only that agents become more likely to select a card as its expected information gain increases. Their revised model follows Hattori (2002) in modeling a logistic relationship between the probability of turning a card and its scaled information gain.[13] And that is not what the reason-responsive consequentialist view from Part 2 requires. Agents should select the cards with the highest expected value. They should not merely become more likely to select the cards with highest expected value. Rationality requires agents to actually do what they ought, not merely be more likely to do so.[14]

This finding is far from ideal, but I think it is only to be expected. If we are not Panglossians, we should not expect to find that agents always do what they ought, and we should be suspicious of models and theories which claim that they do. Given the extremely poor performance claimed by logical models of the Wason selection task, it is vindication enough to find that selection tendencies

---

[11] The uncertainty of an agent's beliefs P is standardly measured by Shannon entropy. That is $I(P) = -\Sigma_{X=\{M_t, M_D\}} P(X) log_2(P(X))$, where $I(P)$ is the uncertainty of P.

[12] So for example, writing $I_p(P)$ for the information-gain of turning card p when holding beliefs P, the scaled information-gain of turning p would be $SEI_p(P) = \frac{E[I_p(P)]}{\Sigma_{x=\{p, \neg p, q, \neg q\}} E[I_x(P)]}$.

[13] That is, letting $ST_x(P)$ be the probability of turning card x, we have $ST_x(P) = \frac{1}{1+e^{2.37-9.06 SEI_x(P)}}$.

[14] A complication: this sort of matching behavior can result from rational sampling of information in memory (Icard 2018; Vul et al. 2014). If that is right, we might say that irrational card selection may be caused by rational information usage during metacognitive monitoring and control of reasoning. But it will not make the resulting inquiries any more rational.

track expected values. This is an under-emphasized feature of recent Bayesian approaches to the psychology of reasoning. Not only the information-gain model, but also many other Bayesian models show only that response tendencies track probabilities or expected values, not that agents always make the best judgment or decision. Again, these findings are unfortunate but expected. There is a wide gap between the skeptical claim that human cognition is mostly irrational and the Panglossian claim that human cognition is fully rational. The information-gain model, like other Bayesian models, pushes us well away from the skeptic while keeping us comfortably away from the Panglossian. I will have more to say about this issue in Section 7.6.5.[15]

7.6.3 Deontic selection

So far I have introduced an information-gain model of the Wason selection task on which agents select cards to reduce their uncertainty about the probabilistic relationship between p and q. I argued that this account coincides with a standard consequentialist account of rational evidence-gathering and substantially vindicates observed behavior in abstract versions of the Wason selection task.

At this point, it may seem unclear whether the information-gain model is deeply consequentialist. A good number of nonconsequentialist approaches to rational inquiry could be finessed to yield the result that agents should gather evidence in an abstract setting in ways that yield the greatest expected uncertainty reduction. So too, restricted consequentialist views such as epistemic consequentialism might deliver that result. In this subsection, I show how the information-gain model has been extended to a less abstract version of the Wason selection task, the *deontic selection task*. Here, I argue that the model imposes normatively correct requirements which substantially vindicate observed behavior. More importantly, the information-gain model applied to deontic selection tasks is unmistakably consequentialist.

In deontic selection tasks, the card labels are replaced with more realistic contents and conditionals are altered to express deontic rules. Studies of deontic selection tasks reveal at least three novel findings. First, card selection depends on whether the rules express permissions (if p, may q), or obligations (if p, must q). With obligation rules, for example:

If a man eats cassava root, then he must have a tattoo on his face.

---

[15] As we will see in Section 7.6, this explanation leaves room for probabilistic reasoning to count as irrational when agents have good reason to be performing logical reasoning instead, or when they engage in incorrect probabilistic reasoning. It is also importantly distinct from the dual-process account on which syllogistic reasoning becomes irrational due to the distorting influence of System 1 thinking (Evans et al. 2003; Evans and Stanovich 2013).

Participants typically turn that p and $\neg q$ cards, checking for cheaters who eat cassava root but don't have facial tattoos (Cosmides 1989). But when given the corresponding permission:

If a man eats cassava root, then he may have a tattoo on his face.

Participants turn the $\neg p$ and q cards, checking for tattooed cassava-avoiders.

Second, deontic selection behavior shows an effect of perspective. Suppose that a mother makes a rule for her child:

If you tidy your room, then you may go out to play.

Participants asked to test the rule from the perspective of the child will turn the p and $\neg q$ cards, checking that children who have tidied their rooms have been allowed to play. But asked to take the mother's perspective, they will turn the $\neg p$ and q cards, checking that children who play have cleaned their rooms.

Third, deontic selection behavior shows sensitivity to the costs of hits and false alarms. Suppose you are a bouncer instructed to test the rule:

If someone drinks, then they must be at least 21 years old.

has been followed. Participants instructed that management wants to avoid hefty fines for underage drinking become more likely to turn each relevant card, avoiding false negatives (Kirby 1994). Participants told that management wants to avoid alienating customers become less likely to turn each relevant card, avoiding false positives. How can these three findings be explained?

Many models of deontic selection behavior are possible. In the 1980s, two accounts were developed which propose that agents have domain-specific methods for solving deontic selection tasks. These can be innate modules for detecting cheaters and dealing with hazards (Cosmides 1989) or learned schemata for reasoning with permissions and obligations (Cheng and Holyoak 1985). By contrast, the more parsimonious information-gain model proposes that agents apply domain-general forms of reasoning in a way that incorporates the particular demands of deontic selection tasks. Exactly the same model can be applied to deontic selection as to the abstract Wason selection task.

In deontic selection tasks, agents are asked to adopt particular perspectives from which the utilities of outcomes are well-defined. For example, a bouncer gains from detecting underage drinkers by preventing potential fines and loses in checking legal drinkers by angering customers. Here the expected value of information-gathering is no longer the reduction of uncertainty, as was appropriate when actions and utilities were left unspecific, but rather these immediate costs and

benefits specified by the problem and the role of the actor.[16] The normative response can be determined by the information-gain model together with the probabilities and utilities drawn from the problem description. This model fits observed behavior on deontic selection tasks reasonably well (Oaksford and Chater 2007; Perham and Oaksford 2005). Differences between rule types combine with perspectives to determine which outcomes matter, and manipulations of costs and benefits determine how much they matter.

In broad outline, the information-gain model explains sensitivity to costs of hits and false alarms by building these costs into the value of information-gathering. As it becomes more costly to alienate customers, the expected value of checking for identification decreases and hence bouncers should be less likely to check for information. The information-gain model explains perspective effects by letting perspectives shift the relevant value function. Children check for cheating parents because it is most important to children that they be allowed to play, whereas parents check for cheating children because it is most important to parents that children clean their rooms and follow household rules. Differences between permissions and obligations are explained in a similar way. If cassava eaters must be tattooed, we check for uninked cassava eaters because it is important to detect theft from the communal larder. If tattooed individuals may eat cassava, we check for tattooed cassava-avoiders because it is important to make sure nobody is being denied their fair share.

Notice that this version of the information-gain model is explicitly consequentialist.[17] It is a straightforward application of expected utility theory, like the Lieder and Griffiths model of anchoring and adjustment in Section 7.5. Traditional epistemological models will have to posit two separate types of normative requirements governing abstract and deontic selection tasks, respectively. They will hold that the normative requirements in deontic selection tasks are very different in kind from requirements in abstract tasks, for example, because the normative requirements in deontic selection tasks are non-epistemic. This separation becomes difficult to maintain when we observe that humans are almost never confronted with purely abstract selection tasks. This means that the type of purely epistemic normativity associated with uncertainty-reduction that is meant to be derived from standard

---

[16] For example, letting p be the proposition that an agent drinks beer, and q be the proposition that an agent is at least 21 years old, the expected value of turning card p is $E[p] = P(q|p)u(pq) + P(\bar{q}|p)u(p\bar{q})$.

[17] Note that while this treatment agrees with some versions of the Standard Picture in its acceptance of expected-utility maximization as a normative standard, it differs from the Standard Picture in many aspects, such as its attention to cognitive architecture and task environments, and its rejection of coherence-based normative standards. It also comes apart from the Standard Picture in subtler ways, such as its avoidance of axiomatic treatments of rationality (Gigerenzer 2019) and because leading Bayesian approaches to the psychology of reasoning often bottom out in cognitive heuristics (Oaksford and Chater 2007). For a full discussion of how my view relates to the Standard Picture, see Chapter 9. Thanks to a referee for pressing me to clarify the divergence here.

Wason tasks will almost never govern actual situations of human judgment and decisionmaking. In more complicated situations, we will need a mixed model which incorporates the intrinsic and instrumental value of uncertainty reduction together with the immediate costs and benefits of evidence-gathering actions. The consequentialist model is the best model that I am aware of which is up to the task.[18]

Summing up, the information-gain model can substantially vindicate observed behavior on abstract and deontic versions of the Wason selection task. This model is an extension of the standard consequentialist account of duties to gather evidence. But is it true? In the next subsection, I review evidence for the information-gain model.

7.6.4 Evidence for the information-gain model

What evidence can be given for the information-gain model?[19] The literature on conditional reasoning is extremely large, and it would be impossible to survey even a modest portion of the literature on Wason selection in this subsection, let alone other tasks. Moreover, some accounts of Wason selection behavior are meant to account for many other types of human reasoning beyond conditional reasoning, so to a large extent, the fortunes of accounts may rise and fall with considerations outside the Wason selection data. Nevertheless, it is appropriate to take a brief look at the descriptive evidence for the information-gain model. For a complete survey of evidence for the information-gain model of Wason selection tasks, the reader is referred to Oaksford and Chater (2007). In this subsection, I briefly review three strands of evidence for the information-gain model.

First, probability manipulations show effects on all versions of the Wason selection task, and these effects are broadly consistent with the information-gain model (Kirby 1994; Oaksford and Chater 2003; Oaksford et al. 1999). With the exception of some close cousins of the information-gain model, most competing models predict no effect of probability manipulations whatsoever and the rest predict only much more limited and localized influences.

Second, the information-gain model generalizes to predict results on novel tasks. For example, in the *reduced array selection task* (RAST) participants are only given the option to select between q and $\neg q$ cards. The information-gain

---

[18] We saw in Chapter 5 that some related views, such as Stich-style pragmatism (Stich 1990) can also handle mixed cases, but that there may be grounds to prefer consequentialism to pragmatism. Can other approaches, such as theories of the weight of reasons, handle such cases? Perhaps. It would be an interesting project to develop the consequences of alternative philosophical accounts for cases such as deontic selection and anchoring and adjustment, then compare the results for specificity and plausibility with the results of the RRCV.

[19] For pushback see Vindrola and Crupi (forthcoming) and Oberauer et al. (1999).

model predicts that participants will select mostly q cards when q is improbable, and mostly $\neg q$ cards as q becomes more probable. RAST experiments confirm this prediction and yield generally adequate fits to the information-gain model (Oaksford et al. 1997).

Third, the information-gain model predicts the influence of negations on selection behavior. The *negations paradigm* is identical to the standard Wason selection task, except that one or both terms in the conditional may be negated. In papers discussing the negation paradigm, the convention is to describe conditionals by whether each term is affirmative (A) or negated (N), so that for example the conditional if p, then $\neg q$ would be denoted AN. Cards are described by their relationship to the antecedent and consequent. In this example, p would be the true antecedent (TA) card, and $\neg p$ would be the false antecedent (FA) cards. Similarly, q and $\neg q$ would be the true consequent (TC) and false consequent (FC) cards.

There are several puzzling findings in the negations paradigm. Chief among them is a type of matching behavior. For rules with affirmative consequents (AA and NA), the likelihood of selecting each card follows the order $TA > TC > FC > FA$. For rules with negated consequents (AN and NN), card selection follows the order $TA > FC > TC > FA$. Here it appears that participants select between q and $\neg q$ cards to match the consequent of the rule. If the rule has consequent q, then participants are more likely to select that q card than the $\neg q$ card, and if the rule has consequent $\neg q$, participants are more likely to select the $\neg q$ card. This finding was traditionally explained as a matching bias: participants prefer to select the cards named in the rule (Evans and Lynch 1973). This explanation persisted even when it was discovered that, as the above choice ordering suggests, matching occurs only for the consequent card and not for the antecedent.

The information-gain model can explain this tendency by invoking expected information gains, without positing a new matching bias localized to the consequent of conditionals. The missing ingredient is a *contrast-class account* of negation. On this account, propositions are contextually associated to a contrast class of relevant alternatives, and the negation of a proposition expresses the disjunction of the claims in the contrast class. For example, if p is the claim that a woman at a baseball game is eating peanuts, the contrast class might be the claims that she is eating a hot dog, pretzels, or cotton candy. A consequence of this account, which was developed elsewhere to account for reasoning with negations, is that the negation of an improbable claim r should be more probable than the original claim, but should have probability less than $1 - P(r)$. Oaksford and Chater (1994, 2007) use the contrast-class account of negation to sample a reasonable range of probability assignments on which the information-gain model predicts matching behavior and other data from the negations paradigm without invoking novel psychological mechanisms or biases.

Summing up, while it was not possible to survey the vast literature on Wason selection tasks, we have seen three strands of evidence in favor of the

information-gain model. First, the model correctly predicts the existence and effects of probability manipulations. Second, the model generalizes to new tasks such as the RAST. Third, the information-gain model predicts performance in the negation paradigm. The reader is referred to Oaksford and Chater (2007) for further evidence for the information-gain model.

7.6.5 Banishing Pangloss

I want to conclude on a more somber note. While the information-gain model goes a long way toward vindicating observed behavior on the selection task, it also reveals some clear irrationalities. I will discuss four examples in this subsection.

First, as we saw in Section 7.6.2, the information-gain model predicts a logistic relationship between scaled expected information gains and selection behavior. By contrast, consequentialists hold that the normatively correct behavior would be to always choose the cards with the highest expected information gain. Moreover, the information-gain model often improves model fits by assuming a slight probability that participants will choose randomly. Random choice is not normatively correct behavior.[20] While this is unfortunate, it is a very common finding in choice experiments that model fits are improved by assuming a small probability of random choice. Hence I do not think that this finding reveals any special problem for choice behavior on the selection task.

Second, sometimes the information-gain model fits data less than perfectly and in at least one dramatic example it fits quite poorly indeed (Oberauer et al. 1999). Without entering too deeply into descriptive debates, I want to suggest that even if some of these difficulties can be chalked up to methodological differences, part of the problem is that agents do not always choose normatively. When there are persistent difficulties fitting normative models to data, the most natural conclusion is that some of the choices in the data are not normative.

Third, several applications of the information-gain model improve model fits by using probability-weighting. Tversky and Kahneman's prospect theory introduced the concept of a probability weighting function π, which weights probabilities in expected utility calculations, just as a utility function u weights outcomes.[21] Prospect theory suggests that all human decisionmaking involves probability weighting, hence this finding should not be particularly surprising. Nevertheless, on my view probability weighing is probably irrational. Axiologically, the value of an action is its unweighted expected utility, not its probability-weighted expected utility. I am open to the possibility that probability weighting could be a rational decision heuristic, but I don't know of decisive arguments for that conclusion.

---

[20] Or rather, existing justifications for random choice (Icard 2021) will not vindicate this behavior.
[21] That is, we have $EU[g] = \Sigma_w \pi(p(w))u(g(w))$ for gambles g.

Fourth, the probabilistic turn has changed the subject. We asked whether humans are good logical reasoners and answered that experiments purporting to show that humans are bad logical reasoners in fact reveal that we are fairly good probabilistic reasoners. But we can still ask our original question: are humans good logical reasoners? Nothing that has been said in this section suggests that we are good logical reasoners, and in fact that probabilistic turn raises at least one reason to suspect that we are not.[22] Namely, it is held that humans rarely use logical reasoning because probabilistic reasoning is often more appropriate to the problems that we face. If this is right, then it would not be particularly surprising to find that typical humans have neither learned nor evolved the capacity for effective logical reasoning. By contrast, that finding was especially surprising when it was earlier supposed that most human reasoning aims to solve a logical deduction task. Hence the present argument gives some reason to doubt that humans are good logical reasoners, but at the same time it softens the blow. If we are not especially good logical reasoners, that is because it is not especially important for us to be good logical reasoners.

In this section, I have used the Wason selection task as a case study to argue that human logical reasoning has been unfairly criticized. What looks like poor logical reasoning is often good probabilistic or causal reasoning in disguise. I presented an information-gain model on which performance on the Wason selection task is rational evidence-gathering, which on the consequentialist account amounts to maximizing the expected value of gained information. I argued that on this understanding, performance on the Wason selection task is far from perfect, but it is much better than has been supposed.

7.7 Conclusion

My project in this chapter was to show how the reason-responsive consequentialist view grounds one part of the bounded rationality program outlined in Chapter 2: the vindicatory project of recasting seeming irrationalities as the results of rational cognition by bounded agents. I set out four criteria on an account of vindicatory epistemology. A good account should be genuinely vindicatory, and

---

[22] At the same time, it is important not to be too pessimistic about human logical reasoning ability. We saw in Chapter 1 that even Tversky and Kahneman take biases to be rare, so we should not begin theorizing with the assumption that humans will be terrible logical reasoners. Moreover, recent authors have stressed that tracking logical validity may have value for Bayesian agents (Eva and Hartmann 2018), and that much everyday argument is in line with logical laws (Hahn 2020). Similarly, some studies have found that humans are guided by the logical form of arguments as well as by Bayesian rules for incorporating prior knowledge (Singmann and Klauer 2011; Singmann et al. 2016), and even as previous logic-based paradigms in the psychology of reasoning (Evans and Lynch 1973; Rips 1994) lose some of their popularity, we should take their explanatory successes as some evidence that humans often reason well according to logical laws.

the vindications offered should be continuous with existing scientific vindications when these vindications are plausible. Vindicatory programs should also steer comfortably away from the Panglossian claim that human cognition is as rational as it could possibly be, and should capture a conflicted stance of evaluative tension toward agents in many tasks.

I used case studies of conditional reasoning and anchoring and adjustment to show how the RRCV gives rise to a vindicatory program with the desired features. In both cases, we saw that the RRCV can vindicate wide swaths of human cognitive practices. These vindications make direct use of existing consequentialist models in the scientific literature. The program is not Panglossian: we saw in Section 7.6.5 that the vindication of conditional reasoning leaves open four types of irrationality, and we saw in Section 7.5.3 that standard anchoring paradigms may produce a localized form of irrationality in participants' response to the second, absolute question posed to participants. And my vindicatory program captures evaluative tension by taking many of these cases to involve rational inquiry leading to irrational beliefs.

If my arguments in this chapter are successful, they show that the RRCV can do one part of what it was meant to do: ground a program of vindicatory epistemology. But Chapter 2 set out four more claims for an account of rational inquiry to ground and explain. What light can my theory shed on these claims? I take up that question in the next chapter.

8

Bounded rationality revisited

8.1 Introduction

In Chapter 7, we saw how the reason-responsive consequentialist view can carry out one part of the bounded rationality program outlined in Chapter 2: the vindicatory project of recasting seeming irrationalities as in an important sense fully rational. In this chapter, I show how the RRCV accounts for the other four normative claims made in Chapter 2: that bounds matter (Section 8.2), and that rationality is procedural (Section 8.3), heuristic (Section 8.4), and ecological (Section 8.5). In each case, my aim is to do at least three things.

First, I aim to ground each normative claim, explaining not only why the claims are true but also explaining why traditional arguments for each claim are good arguments. For example, we saw in Chapter 2 that there are three traditional arguments for the rationality of heuristic cognition, and in this chapter I will show how the RRCV underwrites each argument. Second, I aim to clarify the contents and consequences of each normative claim. For example, I want to say what it means to claim that rationality is procedural or ecological, and to explain why my interpretations might be normatively preferable to other competing interpretations of these concepts. Finally, I aim to generalize the normative claims made in Chapter 2 together with the arguments for these claims. This will provide explanatory illumination by revealing some traditional claims and concepts to be consequences of more general normative phenomena, and will also suggest new directions for future research.

8.2 Bounds matter

8.2.1 The relevance constraint

Humans are not angels, but creatures of flesh and blood. Just as our bodies constrain the physical actions we can undertake, so too our minds constrain the mental actions we can undertake. For this reason, the bounded tradition holds that bounds matter to rational cognition.

Chapter 2 split the claim that bounds matter into two parts. The first is a relevance constraint:

Relevance constraint: Paradigmatic cognitive bounds including limitations on agents' cognitive abilities as well as the costs of cognition bear on the rationality of agents' judgment and decisionmaking.

The relevance constraint asserts that two paradigmatic bounds are rationally relevant: limited abilities and cognitive costs. The reason-responsive consequentialist view not only grounds but also widens the relevance constraint.

How does the RRCV account for the normative relevance of limited cognitive abilities? We saw in Chapter 3 that the RRCV holds that rationality is deontic (Lord 2017): what agents ought to do and what they are rationally required to do are the same thing. That is:

(Rationality is Deontic) For all agents S, times t and features of agency X, S is rationally required at t to X if and only if S ought at t to X.

In deontic matters, ought implies can:

(Deontic OIC) For all agents S, times t and features of agency X, if S ought at t to X then S can X at t.

Putting these principles together shows that ought implies can for rational requirements as well. That is:

(Rational OIC) For all agents S, times t and features of agency X, if S is rationally required at t to X then S can X at t.

Rational OIC captures the normative relevance of limited abilities: agents cannot be rationally required to do what it is beyond their abilities to do.[1]

The reason-responsive consequentialist view also captures the second part of the relevance constraint: that the costs of computation bear on the rationality of agents' judgment and decisionmaking. At the very least, computational costs are opportunity costs: if we did not spend scarce cognitive resources such as computational bandwidth, attention and working memory space on one task, we could spend these resources on another. If it would be better to engage in that task instead, then the RRCV says that wasting computational resources on a less-fruitful task is irrational because it is not best. The RRCV also allows that computational costs may be relevant in other ways. For example, if you spend time

---

[1] Note that (Rational OIC) implies that agents with limited abilities may be rationally required to engage in processes of reasoning, or hold beliefs, which agents with more robust abilities would be rationally forbidden from engaging in or holding. If a creature's abilities are limited enough, we may want to hold instead that the creature is not sufficiently agential to be governed by rational requirements, or even to be creditable with beliefs, desires, and reasoning processes.

thinking about philosophy while traveling to the beach, you will be mentally tired when you arrive and may not enjoy the beach. Because it is good to arrive rested and refreshed to the beach, under many circumstances agents may be rationally required not to exhaust themselves on the way to the beach.

The RRCV also generalizes the relevance constraint by explaining how noncognitive costs incurred during inquiry bear on the rationality of inquiry. Studying philosophy consumes scarce resources such as time, money, and the cost of carbon during travel to conferences. Both time and money are relevant as opportunity costs since those resources could have been spent on other inquiries or leisure activities. Carbon emissions matter not as opportunity costs, but because they contribute to global warming, which imposes a variety of well-known harms on future populations. In this way, the RRCV avoids a narrow construal on which the only costs that matter are forms of mental effort.[2] Instead, the RRCV holds that all of the costs of inquiry matter rationally.

The reason-responsive consequentialist view generalizes the relevance constraint in another way: by explaining how bounds beyond limited abilities and cognitive costs matter rationally. For example, humans are informationally bounded agents: we have only so much information about the world. We saw in Chapter 3 that objectivist accounts of rational inquiry deny that informational bounds bear on rational cognition. We also saw that subjectivist accounts cannot properly incorporate informational bounds when agents lack beliefs about relevant matters, or when those beliefs do not fully incorporate available information. For that reason, the RRCV adopts an information-sensitive account of rationality which builds the relevance of informational bounds directly into the story about possessed normative reasons, against which the rationality of inquiry is determined.

So far, we have seen how the RRCV grounds and generalizes the relevance constraint. The RRCV explains the rational importance of limited cognitive abilities and the costs of cognition and incorporates other bounds such as noncognitive costs and informational limitations. We will see below that the RRCV also reveals a hidden ambiguity in the relevance constraint and helps us to see the normative importance of this ambiguity.

8.2.2 Relevance to what?

To say that bounds are relevant to rational cognition is not yet to say which aspects of cognition they are relevant to. In particular, we may wonder whether bounds bear on the normative status of attitudes or the normative status of the inquiries that produced them. By implementing a sharp level separation

---

[2] For a review of conceptions of mental effort, see Shenhav et al. (2017).

(Chapter 3) between different evaluative focal points, such as inquiries and the attitudes they produce, the reason-responsive consequentialist view helps us to see the importance of specifying the focal points on which paradigmatic bounds bear.

Because bounded rationality is process focused, I have primarily discussed bounds on cognitive processes, such as our limited abilities to carry out complex computations and the costs incurred by those computations. By level separation, these facts may not have any direct bearing on the rationality of belief. To say, for example, that the calculations needed to produce a belief would be costly is not yet to say that it is costly to hold that belief itself. For that reason, we may hold that it would be wastefully irrational to count the blades of grass on your lawn, but that if you were to do so, your belief about the number of blades of grass could count as rational.

On the RRCV, are bounds such as costs and limited abilities also relevant to the normative status of attitudes? The first thing to say here is that insofar as bounded rationality is process-focused, questions about the rationality of attitudes are not always the most revealing lens into bounded rationality. We can, if we wish, ask how bounds bear on the rationality of attitudes, but the answer to this question may not tell us very much about the overall normative importance of cognitive bounds.

Nevertheless, to ask whether bounds bear on the rationality of attitudes is to ask a well-formed normative question. There is some reason to think that bounds may bear on the rationality of attitudes. For example, beliefs may be beyond our ability to hold if they are too complex or heinous to contemplate. And beliefs may be costly if they are so tragic as to make us sad. What does the RRCV say about the rational importance of costs and limited abilities in such cases?

There are at least two ways to hold that the RRCV says nothing about such cases. First, the RRCV is only a claim about the rationality of features of agency. Many readers may think that attitudes such as beliefs are not features of agency. For example, they may think that features of agency must be voluntary but that beliefs are involuntary. On this account, features of agency are not attitudes such as beliefs, intentions, and preferences, but rather actions such as the processes of inquiry that produce our attitudes. This would make space for combining a consequentialist account of rational inquiry with traditional accounts of rational belief, intention, and preference. Those accounts differ amongst themselves as to whether and how cognitive bounds bear on the rationality of attitudes.

In the special case of rational belief, we could also distinguish between epistemic and all-things-considered notions of rationality, or even deny that it makes sense to talk about rationality simpliciter. Although Chapter 4 argued against this move in the case of rational inquiry, that argument did nothing to challenge the case for epistemic norms governing belief. In fact, we saw in Chapter 4 that there are at least three classes of arguments that could be given for positing a distinct class of epistemic norms governing belief. If this is right, then even if the RRCV has

implications for the all-things-considered importance of bounds on attitudes, we are still free to adopt other accounts of epistemically rational belief, and these accounts may come apart from the RRCV in their treatment of cognitive bounds.

If we do not adopt either of these strategies, then the RRCV will say that costs and limited abilities may sometimes bear on the rationality of belief. If a belief is too complex to even consider, then we will not be rationally required to hold it. And if doubting the existence of God would make us depressed and suicidal, then we may be required to believe in God. But if we think of beliefs, intentions, and other attitudes as sufficiently voluntary to count as features of agency, and if we also do not insist on a separate epistemic reading of norms governing belief, then these verdicts no longer appear implausible. If beliefs are like actions in being voluntary undertakings, then it may make no more sense to say that you should believe a proposition too complex to contemplate than to say that you should lift a skyscraper with your bare hands. And if you could voluntarily believe in God to forestall suicide, then that might very well be a rational thing to do.[3]

Summing up, we have seen that the reason-responsive consequentialist view focuses on the normative relevance of bounds to processes rather than attitudes. On some ways of spelling out the RRCV, it may not pronounce on the rational importance of bounds on attitudes. On other readings, the RRCV will say how bounds impact the rationality of attitudes, but these readings are not implausible. More importantly, they are not the most revealing lens into the normative importance of cognitive bounds.

8.2.3 The methodological constraint

We saw in Chapter 2 that in addition to the relevance constraint, the claim that bounds matter also imposes a second methodological constraint:

> **Methodological constraint:** Relevant bounds should be incorporated, as far as possible, into all stages of rationality modeling.

The methodological constraint is meant to block models on which we say first what would be required of unbounded agents, then adjust those models to determine how bounds agents should cognize.

Because the methodological constraint is not a claim about rationality, it is not the sort of principle that could be grounded or explained by the RRCV.

---

[3] Indeed, most epistemologists grant that you should *get yourself* to believe in God in this case. They have held off from saying that you should believe in God based on the two arguments mentioned above: that believing, unlike getting yourself to believe, is involuntary; and that there is an epistemic reading of rationality on which suicide prevention does not bear on rational belief.

Nevertheless, there are at least two ways in which the discussion from Parts 2–3 sheds light on the methodological constraint.

First, the program of vindicatory epistemology (Chapter 7) can be taken to motivate the methodological constraint. Vindicatory epistemology confronts cases in which unbounded normative models have generated predictions that are importantly incorrect. For example, unbounded models predict that rational agents should show no anchoring bias. Vindicatory epistemology then proceeds to show how we can challenge these normative verdicts by building detailed models of bounds directly into cognitive models. If we specify in detail the processes, such as anchoring and adjustment, available to agents, as well as the costs and benefits of implementing these processes, we see that anchoring bias may fall out of rational processes of anchoring and adjustment aimed at striking an optimal balance between accuracy and effort in cognition.

These are not predictions that can be easily made by using or approximating unbounded models. All of the most natural arguments for the rationality of anchoring bias involve incorporating detailed specifications of cognitive bounds into normative models. Indeed, the historical genesis of anchoring bias and other cognitive biases provides some evidence for the methodological importance of attention to cognitive bounds. Many of the best-known cognitive biases, including anchoring bias, were widely regarded as irrational by those who discovered them. Those who have come to regard these biases as rational have done so in large part by obeying the methodological injunction to build bounds into normative models, and showing that these richer models vindicate the rationality of seemingly irrational biases. Theorists who do not build bounds into models of cognitive biases have largely been tempted to classify these biases as irrational, even if they are nothing of the sort.

Second, although the reason-responsive consequentialist view cannot ground or explain the methodological constraint, the RRCV may be our best hope for complying with the methodological constraint. In both of our case studies from Chapter 7, we saw that the best scientific models of relevant cognitive bounds are explicitly consequentialist. And in Chapter 5, we saw that consequentialism tells a detailed story about how normative factors such as stakes and tradeoffs bear on the rational allocation of scarce cognitive resources. More generally, the ability of consequentialist accounts to precisely weigh a variety of competing costs, benefits and constraints has always been considered a virtue of consequentialist theorizing. As consequentialists continue to construct plausible normative models of the problems that confront bounded decisionmakers, these models will combine to support the RRCV as a way to comply with the methodological constraint.[4]

---

[4] For example, there are promising consequentialist models of rational attention (Sims 2003); rational planning (Callaway et al. 2018); and the formation of good cognitive habits (Kermati et al. 2016).

In this section, we saw how the reason-responsive consequentialist view grounds and explains the claim that bounds matter to rational cognition. The RRCV explains the normative relevance of paradigmatic cognitive bounds, including bounds such as informational limitations that other views may struggle to incorporate. The RRCV helps us to tease apart the normative relevance of bounds to attitudes and processes. And the RRCV may be our best hope for complying with the methodological injunction to construct detailed normative models of cognitive bounds.

8.3 Process focus

8.3.1 Interpreting the procedural turn

Herbert Simon held that a fundamental turn in the study of bounded rationality is the turn from substantive to procedural rationality. We saw in Chapter 2 that the procedural turn can be interpreted in two ways.

On Simon's (1976) *substitute* interpretation, substantive and procedural rationality are conflicting normative assessments of the same evaluative focal point: an agent's attitudes. Theories of substantive rationality evaluate attitudes in their own right, and not in light of the processes which produced them. For example, we might say that an attitude is right just in case it is best. Theories of procedural rationality let attitudes inherit the normative status of the processes which produced them. For example, we might say that an attitude is right if it was produced by the best deliberative process. On a substitute interpretation, the procedural turn is important because it allows normative theories to incorporate bounds on cognitive processes, such as limited computational abilities and the costs of computation.

On my preferred *complementary* interpretation, substantive and procedural rationality ask questions at different evaluative focal points and hence are fully compatible. Theories of substantive rationality ask normative questions about attitudes: what is it rational to believe, intend, or prefer? Theories of procedural rationality ask normative questions about processes of inquiry: how is it rational to deliberate about what to believe, intend or prefer? On a complementary interpretation, the procedural turn is important because normative facts about attitudes do not reflect the full importance of many cognitive bounds, which are felt most strongly as bounds on processes rather than attitudes.

The RRCV recasts the distinction between substitute and complementary interpretations of the procedural turn in terms of the familiar distinction between indirect and direct normative assessment. Indirect normative theories assess evaluative focal points such as attitudes at least partly in light of other evaluative focal points, such as processes of inquiry. Simon's substitute interpretation is a fully indirect normative theory, because it lets the normative status of attitudes be

entirely determined by the normative status of the inquiries which produced them. By contrast, direct normative theories assess evaluative focal points in their own right, and not in light of other focal points. The complementary interpretation becomes a direct normative theory when coupled with a strict level separation between normative questions about attitudes and normative questions about the processes that produced them.

A complementary interpretation of the procedural turn is made possible by the common distinction between the normative status of decision procedures and the normative status of the intentions that they produce. The complementary interpretation extends this distinction to other attitudes, distinguishing for example between the rationality of a belief and the rationality of the process of inquiry that produced it. Rephrasing the complementary interpretation in this way reveals at least two motivations for adopting it. First, the distinction between norms governing decision processes and norms governing attitudes is one of the most well-known and fruitful distinctions in modern ethical theory. In this way, the complementary interpretation gains plausibility from its explanatory applications. Second, a sharp distinction between norms governing processes and attitudes is widely regarded as an essential component of many of the most plausible contemporary consequentialist theories (Driver 2012; Parfit 1984; Railton 1984). To this extent, the complementary interpretation gains plausibility from existing arguments for consequentialism.

So far, we have seen that the reason-responsive consequentialist view recasts the debate between complementary and substitute interpretations of the procedural turn in familiar terms, and provides grounds for adopting a substitute interpretation. The next question to ask is why we should take the procedural turn in the first place.

8.3.2 Motivating the procedural turn

On a complementary interpretation, the procedural turn amounts to the claim that normative theories should be process-focused, putting at least as much emphasis on normative questions about processes as on normative questions about attitudes. The RRCV helps us to see why normative theories should be process focused by revealing just how little the rationality of attitudes such as belief and intention has to do with paradigmatic cognitive bounds.

By way of illustration, suppose you are at the supermarket aiming to buy a bottle of balsamic vinegar. Having chosen a particularly exclusive supermarket, you are faced with a long shelf full of vinegars. Which bottle should you buy or intend to buy? For a consequentialist, the answers to these questions are deceptively simple: you should both buy and intend to buy whatever bottle it would be best to buy. This answer has almost nothing to do with your cognitive capacities. It is, with

one exception, the same answer that we would make to an angel. On the reason-responsive consequentialist view, there is one relevant way in which your humanity influences what is rational for you: you have limited information. If you have not read the labels on some bottles or do not have much background knowledge about vinegar, then these informational limitations can affect what it is rational for you to intend. But you may no more protest that it is costly, tiring or time-consuming to calculate the best vinegar purchase than an angel could lodge this protest. On the RRCV, the fact that it would take twenty minutes to calculate the best vinegar purchase is simply irrelevant to the fact that you ought to buy, and intend to buy the very best bottle of vinegar.

We recover the normative relevance of paradigmatic cognitive bounds when we ask a procedural question: how should you decide which bottle of vinegar to buy? Here, your limited computational capacities may make it rational for you to decide heuristically, for example by satisficing. You might fix the following aspirations: buying a bottle of vinegar that is in-budget, unexpired, and certified to come from Modena. You would then search bottles on the shelf one at a time until you found a bottle that meets your aspirations. Then you would halt decisionmaking with the intention to buy that bottle and go on to buy it. If you deliberate in this way, then your resulting intention and action will probably be irrational. You are unlikely to buy or intend to buy the best bottle of vinegar. But in proper perspective, these claims can be seen for what they are: misleading, uncharitable and incomplete normative descriptions of a case in which a bounded agent has gone about choosing a vinegar in a fully rational way, responding to bounds such as the costs of computation which are simply omitted from theories of rational intention.

To say that normative theories should be process focused is to say that insofar as we are interested in capturing the full normative impact of cognitive bounds, we need to set out and emphasize normative facts about deliberative processes and place less emphasis on facts about attitudes. Normative questions about attitudes are not ill-formed, and in some circumstances they may be useful. But on their own, these questions tell an incomplete and potentially misleading story about the life of a bounded agent.

8.3.3 Explaining the disagreement

It is worth pausing to ask why so many theorists in the bounded tradition followed Simon not only in adopting a substitute interpretation of the procedural turn, but also in seeming not to notice the possibility of an alternative complementary interpretation. Why is it that not only the correctness but also the bare possibility of a different interpretation was missed? Here is my best explanation for this divergence.

José Bermúdez (2009) has argued that it is important to sharply distinguish between descriptive and normative theories of rational cognition.[5] In descriptive application, rationality postulates allow us to efficiently characterize an agent's actions and attitudes, as well as to elicit or infer these attitudes based on her verbal and nonverbal behavior. By contrast, normative theories of rationality take for granted descriptive characterizations of an agent's actions and attitudes, then evaluate the quality of her actions and attitudes. Where descriptive theories aim to say how things are with an agent, normative theories aim to say whether the way things are is for good or for ill.

In general, theories of bounded rationality aim to bring descriptive and normative theorizing closer together, not further apart.[6] If bounds matter normatively, then insofar as humans are intelligent creatures, it is likely that as a descriptive matter we do respond to the bounds that we have. The program of vindicatory epistemology goes further still, holding that we not only respond to our bounds, but also often respond to those bounds in something approaching a normatively correct fashion. If that is right, then normative theorizing should often be a good guide to descriptive theorizing. For this reason, I maintained in Chapter 1 that although the topic of this book is normative rather than descriptive, much of my discussion could be applied to the descriptive case as well.

However, level separation provides a clear instance in which descriptive and normative theories of rationality must drift apart, not only in degree but also in kind. In descriptive application, it makes sense to ask questions about rational belief. It also makes sense to ask questions about rational inquiry. But if we have enough information to tell a detailed descriptive story about inquiry, it makes little sense to tell a separate and partially incompatible story about rational belief. Once we know how an agent will inquire, we know what she will believe. If we knew nothing about the processes shaping an agent's beliefs, it might make sense to assume that she holds the beliefs which are best in their own right. But if we can already model her as taking the inquiries that are best in their own right, then we will get a better description of the agent's beliefs by assuming that she holds the beliefs which result from those inquiries. There is no benefit to splitting apart our descriptive models of inquiry and belief.

In advocating a procedural turn, Simon argued that we need to build information about processes into our descriptive models to get accurate descriptions of the attitudes that agents will form. Simon urged us to learn enough about human processes of inquiry to allow us to give more accurate descriptions of human attitudes, allowing us to predict and explain deviations from Standard

---

[5] Bermúdez also distinguishes between two normative projects, one guidance-giving and the other not. This is not a distinction that I make here.

[6] This optimism that the descriptive and the normative will often coincide is one of many surprising commonalities between traditional theories of heuristic rationality and contemporary Bayesian theories of rational analysis (Anderson 1990; Chater and Oaksford 1999b).

Picture models by building bounds and processes into descriptive models. Insofar as Simon and many of the bounded rationality theorists who followed him were primarily interested in descriptive theorizing, they urged us to combine rather than separate our accounts of attitudes and inquiries. That decision was, we have seen, quite correct: in descriptive application, there is little point to level separation.

The topic of this book is normative rather than descriptive. In normative theorizing, there is a good deal to be said in favor of level separation. For example, we will see in Chapter 9 that without level separation, the RRCV would have the revisionary implication that nearly all traditional normative theories of belief, intention, and preference are false. Level separation allows us to block these normative consequences without pushing back against the descriptive importance of building procedural information into social scientific models.

In this section, we have seen how the reason-responsive consequentialist view distinguishes between complementary and substitute interpretations of the procedural turn and motivates a complementary interpretation in normative theorizing. The RRCV explains the need for the procedural turn by revealing just how little impact paradigmatic cognitive bounds have on normative facts about attitudes. And the RRCV makes space for a substitute interpretation of the procedural turn in descriptive theorizing since the strict level separation needed to motivate a complementary interpretation is a move within normative rather than descriptive theorizing.

8.4 Heuristic rationality

8.4.1 Motivations for heuristic rationality

Many approaches to bounded rationality hold that it is often rational for agents to make judgments and decisions using a toolbox of fast-and-frugal heuristics. Chapter 2 gave three arguments for the claim that heuristic cognition is often rational and held that a good normative theory should ground and explain these arguments. In this section, I show how the reason-responsive consequentialist view recovers and deepens the arguments for heuristic rationality. In fact, we will see that much of the work has already been done.

The first reason why heuristic cognition can be rational is that there is often an accuracy-effort tradeoff in cognition (Johnson and Payne 1985). In expectation, producing more accurate judgments and higher-quality decisions requires more effort. Heuristics often strike a good balance between effort and accuracy in cognition, returning highly accurate results at a fraction of the usual cost.

The RRCV explains why the accuracy-effort tradeoff is a good argument for the rationality of heuristic cognition. We saw in Section 8.2 that the RRCV explains

the relevance of paradigmatic cognitive bounds including the cognitive costs of cognition. Cognitive effort matters because it is an opportunity cost: effort spent on one inquiry could instead have been spent on another. Cognitive effort may also matter in its own right, for example because too much thinking makes you tired. We saw that the RRCV folds this account of the importance of cognitive effort into a more general account of the importance of cognitive costs and tells a powerful story about how the costs and benefits of cognition are to be weighed. The accuracy-effort tradeoff falls out as a special case in which the benefits of accuracy must be weighed against the costs of cognitive effort spent to attain it.

More generally, the lives of bounded agents are fraught with tradeoffs. The reason-responsive consequentialist view helps us to see how many of these tradeoffs may tell in favor of heuristic cognition.[7] For example, there is often a *speed-accuracy tradeoff* between the speed and accuracy of judgment and decisionmaking (Heitz 2014). If a lifeguard sees three drowning swimmers, she can make a better decision about whom to save first if she deliberates for thirty seconds rather than one second. But the cost would be a thirty-second delay that even the best rescue plan is unlikely to compensate for. A simple heuristic such as saving the nearest swimmer first would perform better. In situations such as these, heuristic cognition may be rational when heuristics provide results that are reasonably accurate and much quicker than the results of nonheuristic cognition. The RRCV explains why the speed-accuracy tradeoff is a good argument for heuristic cognition. The expected consequences of deliberation include the time spent deliberating, and if that time could have been better spent swimming, then we are rationally required to deliberate quickly and start swimming.

A second argument for the rationality of heuristic cognition cites our limited cognitive abilities. Sometimes humans are unable to execute complex nonheursitic processes at any cost.[8] In that case, the argument continues, we are not required to cognize nonheuristically because we cannot do so. We saw in Section 8.2 that the RRCV grounds the relevance of limited abilities by generalizing the principle that ought implies can from oughts to rational requirements:

(Rational OIC) For all agents S, times t and features of agency X, if S is rationally required at t to X then S can X at t.

By Rational OIC, if agents are unable to employ some complex nonheuristic method, then they cannot be rationally required to do so. Importantly, the reason-responsive consequentialist view does not take Rational OIC to be a brute, unexplained principle. Rather, the RRCV derives Rational OIC from a

---

[7] In Thorstad (forthcoming), I argue that a tradeoff between accuracy and coherence also tells in favor of heuristic cognition.

[8] For example, in full generality many forms of Bayesian reasoning (Chickering et al. 2004; Shimony 1994) well as their approximations (Abdelbar and Hedetniemi 1998; Dagum and Luby 1993; Kwisthout et al. 2011) are NP-hard, a standard measure of extreme complexity.

reason-responsiveness account of rationality together with the popular deontic principle that ought implies can. In this way, the RRCV explains the relevance of limited abilities to the rationality of heuristic cognition by deriving this relevance from prior principles.

The third argument for the rationality of heuristic cognition draws on less-is-more effects (Geman et al. 1992; Gigerenzer and Brighton 2009). In some situations, heuristics reliably outperform more complex nonheuristic methods by avoiding overfitting. When that is the case, the RRCV holds that heuristic cognition is often rational because it provides, in expectation, a greater benefit at a lower cost.

An advantage of this approach to less-is-more effects is that it explains why heuristic cognition could sometimes be irrational, even if it is more accurate and efficient than competing methods. For example, in preparation for standardized tests, students are often taught sets of problem-specific heuristics for simplifying or solving types of mathematical problems that are likely to occur on the test. Learned well, these heuristics may be more accurate and efficient than slower and more error-prone forms of nonheuristic reasoning. But now suppose that a student is practicing for a non-standardized mathematics exam and notices that her heuristics could be applied to a practice problem for the exam. Even though these heuristics are likely to be more accurate and less effortful than other available strategies, we would like to say that the student should not apply heuristics in this case, but should reason explicitly. Why is that?

A natural suggestion is that the aim of solving practice problems is to promote understanding of the underlying mathematical concepts, and that special-case heuristics hamper rather than promote understanding. In contexts such as this one, understanding takes precedence over competing goals such as accuracy and efficiency, hence it would be irrational to cognize using the heuristics in question. The reason-responsive consequentialist view delivers this explanation while still allowing us to hold that in other contexts, such as when taking a standardized test, it may be rational to apply the same heuristics to similar or identical problems. Because the RRCV does not make the intellectualist claim that knowledge and understanding are always the primary goals of inquiry, it allows us to give competing goals such as accuracy and effort their due without denying that understanding has value, or that agents should sometimes pursue understanding.

8.4.2 Strategy selection

Chapter 2 held that the fundamental problem of heuristic cognition is strategy selection. Bounded agents need to choose among a wide array of heuristic and nonheuristic methods as well as to set the internal parameters of these methods. What does the RRCV have to say about strategy selection?

Here it is important to distinguish between two different things that may be meant by strategy selection. On an inquiry-centric reading, questions about strategy selection are questions about inquiry: which strategies should agents use during inquiry, and how should those strategies be parameterized? On a higher-order reading, questions about strategy selection are not questions about inquiry at all, but rather questions about the metacognitive processes by which inquiries are selected (Lieder and Griffiths 2017; Marewski and Schooler 2011). Through metacognitive monitoring of inquiry, agents use beliefs and feelings to represent features of inquiry to themselves. Through metacognitive control, costly control signals are sent to redirect inquiries when they go off course. On a higher-order reading, questions about strategy selection are metacognitive questions: how should agents monitor and control the strategies used during inquiry?

On the reason-responsive consequentialist view, it would be a mistake to run together normative questions about metacognitive strategy selection with normative questions about the processes of inquiry that result, just like it would be a mistake to run together normative questions about inquiry and the beliefs that result. If metacognitive processes are features of agency, then the RRCV will have implications for rational metacognition.[9] But now is not the time to develop those implications. In this section, my concern is with inquiry-centric questions about strategy selection: of the many processes of inquiry available to us, which should we choose?

By way of illustration, Chapter 2 gave three dimensions along which heuristic strategies may vary. First, strategies vary in their cue utilization: the number of decision cues or items of information used to make a judgment or decision. At one end of the spectrum, one-reason heuristics use only a single item of information to make judgments or decisions, whereas other heuristics may use much or all of the information available. Second, strategies vary in their compensatoriness: the degree to which shortfalls along one cue can be compensated by stellar performance along others. Some heuristics, such as satisficing, are maximally non-compensatory, allowing no tradeoffs of any kind, whereas other heuristics may be highly compensatory. Finally, strategies vary in their disposition to incoherence. Although most heuristics are disposed to produce incoherent results in some situation or another, heuristics may vary widely in their likelihood of producing incoherent judgments and decisions in the situations that humans actually face. Chapter 2 held that a good normative theory should shed light on strategy selection by helping us to see how factors such as cue utilization, compensatoriness

---

[9] For an example of a consequentialist account of rational metacognition, see Lieder and Griffiths (2017). As stated, that account may run some danger of running together metacognitive questions with inquiry-centric questions, but something like this account could probably be repurposed as a higher-order account.

and disposition to incoherence are to be balanced against one another. What can the reason-responsive consequentialist view tell us about strategy selection?

On the RRCV, both compensatoriness and cue utilization are red herrings. They matter not in themselves, but rather only to the degree that they affect the expected results of heuristic cognition such as judgmental accuracy or cognitive costs. In many situations, increases in compensatoriness and cue utilization may increase both accuracy and cognitive costs, in which case the right degree of compensatoriness and cue utilization is governed by the accuracy-effort tradeoff. But as we have seen, there is not always an accuracy-effort tradeoff in cognition, and accuracy and effort are far from the only features of cognition which matter.

More generally, this discussion reveals a modest anti-intellectualist undertone to the RRCV. On the reason-responsive consequentialist view, the internal structure of heuristic strategies, such as their cue utilization and compensatoriness, matters not in its own right, but only insofar as it affects the likely results of heuristic cognition. In some cases, it can be rationally permissible to use strategies that radically under-represent or mis-represent reality, so long as those strategies are expected to produce good results. For example, we saw that it may be rational to choose products at the supermarket by the recognition heuristic of buying only products that you recognize, even though many features of a product besides its name-recognition contribute to the overall quality of the product (Goldstein and Gigerenzer 1999). As a consequentialist theory, the RRCV asks us to evaluate heuristics such as recognition not by their internal structure, but rather by the cognitive and noncognitive costs and benefits we are likely to see from using them.

On the RRCV, the disposition of heuristic strategies to produce incoherent results is a very different matter. We saw in Chapter 1 that incoherence is often instrumentally bad, reducing accuracy and creating opportunities for agents to be exploited. We may also think that incoherence is an intrinsically bad state for agents to find themselves in.[10] However, we saw in Chapter 5 that coherence may come apart from other goals, such as accuracy or welfare, that we also have reason to promote. When this happens, the RRCV tells us to balance the intrinsic and instrumental value of coherence against other cognitive and noncognitive goals. It can be rational to cognize heuristically, even when we could use more coherent nonheuristic processes instead, in order to promote other goals such as accuracy, speed, efficiency, or well-being.

In this section, we have seen how the reason-responsive consequentialist view accounts for the rationality of heuristic cognition. The RRCV grounds arguments for the rationality of heuristic cognition based on the accuracy-effort trade-off, limited cognitive abilities, and less-is-more effects. It also generalizes those

---

[10] Some philosophers hold that coherence has no value on its own (Kolodny 2005). Many advocates of coherence requirements hold that coherence has some small intrinsic value, but this value can be outweighed (Broome 2013).

arguments, incorporating new tradeoffs such as the speed-accuracy tradeoff and explaining why even less-is-more effects may not always be sufficient to rationalize heuristic cognition. We also saw how the RRCV can be used to answer specific questions about the rationality of heuristic strategy selection. The RRCV takes a modest anti-intellectualist stance on which the internal structure of cognitive heuristics, such as their compensatoriness or cue utilization, matters only insofar as this structure impacts the likely results of cognition. By contrast, other features such as the disposition to produce incoherent results matter directly and must be balanced together to determine what rationality requires.

8.5 Ecological rationality

The bounded tradition holds that rationality is *ecological*, or environment-relative. Because heuristic strategies perform well in some environments and poorly in others, we cannot ask of a heuristic whether it is rational or irrational full-stop. We must always ask: in which environments would this strategy be a rational way to cognize? The reason-responsive consequentialist view captures the basic datum that rationality is ecological without a hitch. Environments structure the strategies that are available to us as well as the costs and benefits of using them. We have seen that all of these factors are normatively relevant on the RRCV. Strategies that are unavailable to us cannot be rationally required, and we must choose among available strategies based on their expected costs and benefits.

Chapter 2 suggested that a good account of ecological rationality should answer three questions. First, while it is widely agreed that cognition is ecologically rational when it fits the environment in some relevant way, there is no consensus about what type of fit between strategy and environment is at issue. I suggested that the relevant type of fit could neither be fittingness, in the philosophers' normative sense, nor resemblance between strategies and environments. And I suggested that a good first pass was to say that strategies are ecologically rational when we *ought* to employ them in a given environment.

Now we are in a position to improve on that first pass. Ecological rationality is in the first instance a view about *rationality*, not rightness. How could it be otherwise? Strategies are ecologically rational when they are rational to use in a given environment. We saw above that rationality is deontic: what agents ought to do and what they are rationally required to do are the same thing. Because rationality is deontic, the question of how agents are rationally required to cognize in an environment is coextensional with the question of how they ought to cognize. But in general, we saw in Chapter 3 that rationality is a strictly more demanding status than rightness, and it is important to keep rationality and rightness apart.

The first question about ecological rationality gave rise to a second: what is the right deontic theory to think about ecological rationality? I have argued that the

reason-responsive consequentialist view is a good candidate, not only because it induces the right deontic theory, but more generally because it delivers the right theory of rational cognition for bounded agents. At the end of this section, we will have seen that the RRCV does everything that we demanded of a normative theory of bounded rationality: holding that rationality is bounded rationality; capturing the procedural turn; explaining the rationality of heuristic cognition; making rationality ecological; and supporting a vindicatory program. We have also seen that the RRCV is our best hope for meeting three minimal criteria on an account of bounded rationality (Chapter 5) and that the RRCV unifies and explains a range of data from the epistemology of inquiry (Chapter 6).

Our last question about ecological rationality was whether ecological rationality concerns process- and environment-types, tokens, or both? Reliabilists and many other externalists have long held that we need to evaluate types rather than tokens to get plausible verdicts in cases of interest. For example, suppose I ask you when George Washington was born. Here is an eminently efficient and reliable process token for resolving this problem: immediately output the answer '1732' and halt deliberation. On token forms of reliabilism and many other externalist views, that strategy would be rationally permissible, possibly even rationally required.

The standard externalist solution to such worries is to think about process- and environment types. Although the process token of immediately outputting '1732' is reliable in this token environment, it would not be reliable across a type of similar environments in which, for example, questions about the birthdays of other historical figures were posed. Likewise, some reliabilists have held that the relevant process may not be something highly specific such as outputting '1732,' but something more general such as randomly guessing, and this process type is unreliable in most environments. In this way, reliabilists and many other externalists salvage their accounts by turning from tokens to types.

The reason-responsive consequentialist view might perhaps have some resources for resisting this move.[11] But at the same time, the reliabilist has a point. Sometimes, assessing process- and environment tokens just does not tell us what we want to know. On the RRCV, if the agent's evidence supports the judgment that George Washington was born in 1732, then the process token of immediately outputting '1732' will probably be best. Perhaps there is a sense in which agents are rationally required to use that strategy, but that is not the sense we are usually after in asking questions about rational inquiry.

---

[11] Because rationality is information-sensitive, the reliability of a strategy is not a function of its actual accuracy, but rather its expected accuracy given relevant information. An agent whose evidence did not indicate that George Washington was born in 1732 might be irrational for using such a strategy. Likewise, the rationality of a strategy is not a function of the answer that it actually outputs, but rather the agent's evidence about what its outputs will be. In this example, it is clear what the strategy will output, but most strategies involve steps like sampling memory whose consequences are unknown to the agent. This creates room to say that strategies whose answers will in fact be highly accurate may, given the evidence, be counted as unreliable.

One benefit of the consequentialist framework of evaluative focal points is that it allows us to ask several different normative questions about inquiry. On this framework, there are many different focal points: process types, process tokens, belief types, belief tokens, character types, character tokens, and the like. At each focal point, we can ask normative questions about rationality, blame, or other normative categories. Although each question is well-formed, it is not part of the consequentialist view that each question is equally informative or tracks ordinary normative discourse. For example, there is a perfectly well-formed question of which lifetime pattern of behaviors on Tuesday evenings would be rationally permissible. But that is not always a helpful question to ask or a question that most ordinary agents ask.

Likewise, we may agree with the reliabilist that it is often helpful to ask normative questions about process- and environment types rather than tokens, and that this is often what ordinary agents do. We need not go so far as to deny that questions about process- and environment tokens are helpful or well-formed. But we may yet think that an important turn in the study of ecological rationality is the turn toward a greater role for type- over token-assessments.

In addition to answering our original questions about ecological rationality, the reason-responsive consequentialist view helps us to soften some of ecological rationality's surface externalist commitments. We saw in Chapter 3 that a good theory of bounded rationality should say how informational bounds impact rational cognition, and that this desideratum comes into tension with approaches such as externalism. For an externalist, to say that rationality is environment-relative is to say that we are rationally required to use the strategies that perform best in our actual environment, even if all of our best information suggests that the environment will be unfriendly to those strategies. Chapter 3 argued that the best way to soften these commitments is not to adopt a subjectivist approach on which what matters are agents' beliefs about the environment, but rather to take an information-sensitive approach on which what matters is our information about the environment. On this approach, to say that rationality is ecological is to say that agents are required to do what is expectedly best given their information about the environment. An information-sensitive approach allows us to retain the core insights of ecological rationality without denying the normative relevance of informational bounds. Because strategies perform well in some environments and poorly in others, rational agents use strategies that are expected to produce good results given available information about the environment.

Summing up, we have seen that the reason-responsive consequentialist view captures the idea that rationality is ecological and strips this idea of its strongest externalist commitments. The RRCV explains the type of fit between environments and strategies that is at issue in theories of ecological rationality: not resemblance, fittingness, or rightness, but rather rationality. The RRCV provides a promising deontic theory for ecological theorizing and embeds that theory into a more

general account of rationality. And the RRCV accommodates the idea that theories of ecological rationality should often focus on process types rather than process tokens, while at the same time allowing that questions about process tokens are important and well-formed.

8.6 Conclusion

In this chapter, we have seen how the reason-responsive consequentialist view accounts for four characteristic claims about bounded rationality: bounds matter, and rationality is procedural, heuristic, and ecological. We saw that the RRCV grounds each of these normative claims in a natural way, explaining for example why paradigmatic bounds matter and capturing traditional arguments for the rationality of heuristic cognition without change. We also saw that the RRCV can be used to flesh out and develop many characteristic claims about bounded rationality, such as the notion of ecological rationality. And we saw that the RRCV can be used to generalize many of the claims and arguments made in Chapter 2, for instance by revealing the accuracy-effort tradeoff to be one of many different tradeoffs which support the use of heuristic strategies.

At this point, we have done most of what we set out to do. Chapter 2 developed the theory of bounded rationality as a systematic alternative to the Standard Picture and revealed the need for a systematic account of rational inquiry in order to ground and develop the theory of bounded rationality. Parts 2–3 developed and defended the RRCV as an account of rational inquiry. This chapter, together with Chapter 7, showed how the RRCV can be applied to ground and develop all of the claims about bounded rationality made in Chapter 2.

Chapter 1 began by contrasting theories of bounded rationality to the Standard Picture on which rational norms are norms of consistency typified by logic, probability theory, and decision theory. Initially, I suggested that the right theory of bounded rationality would be a normative replacement for the Standard Picture. But now we are in a position to see that this is not quite right. In the next chapter, I use the reason-responsive consequentialist view to revisit the relationship between bounded rationality and the Standard Picture and argue that the two theories are more compatible than they are often thought to be.

9

The Standard Picture revisited

9.1 Introduction

What is the relationship between bounded rationality and the Standard Picture? Chapter 1 suggested that the right theory of bounded rationality will be a normative replacement for the Standard Picture. I suggested that the Standard Picture reacts to descriptive violations of Standard Picture norms by blaming agents, retaining the Standard Picture as a normative theory, and holding that agents who violate Standard Picture norms are cognizing irrationality. By contrast, I suggested that theories of bounded rationality react to many Standard Picture violations by blaming the theory, rejecting the Standard Picture, and replacing it with a competing normative theory on which many Standard Picture violators are cognizing rationally.

Now we are in a position to see that this first-pass description is not quite right. In Section 9.2, I argue that the right theory of bounded rationality charts a third way between blaming the agent and blaming the theory: it blames attitudes rather than agents and blames theorists rather than theories. This allows us to treat theories of bounded rationality not as normative alternatives to the Standard Picture, but rather for the most part as important complements to Standard Picture norms.

That raises the objection that my approach cedes too much ground to the Standard Picture (Section 9.3). I argue that my concessive stance toward the Standard Picture is needed to avoid two normative consequences: that most traditional attitudinal norms are false (Section 9.4), and that all attitudinal norms are normatively non-fundamental (Section 9.5). Section 9.6 concludes.

9.2 A third way

Chapter 1 introduced the Standard Picture of rationality on which rationality consists in conformity to coherence requirements such as the requirements of logic, probability theory and decision theory. We saw that the Standard Picture has been advanced both as a descriptive theory of how humans actually cognize as well as a normative theory of how humans ought to cognize. Chapter 1 chronicled a range of descriptive deviations from Standard Picture axioms and set out two normative reactions that we can take toward these deviations.

First, we can blame the agent, retaining the Standard Picture as a normative theory and holding that agents violate Standard Picture requirements because they are irrational. Alternatively, we can blame the theory, using patterns of descriptive behavior to identify normatively relevant factors which the Standard Picture leaves out, but to which agents are correctly sensitive. Then we can construct an alternative normative theory and use that theory to rationalize many Standard Picture violations.

Initially, I suggested that theorizing about bounded rationality is an exercise in blaming the theory. The right theory of bounded rationality incorporates bounds such as limited cognitive abilities and the costs of cognition to which the Standard Picture is insensitive. As a result, I suggested we might take theories of bounded rationality as normative replacements for the Standard Picture.

But now we are in a position to see that this first suggestion was too strong. The Standard Picture is an account of the rationality of attitudes such as intention, belief, and preference. We saw in Chapters 3 and 8 that the reason-responsive consequentialist view may not have any implications for the rationality of these attitudes, and that if the RRCV does have implications for the rationality of attitudes, these implications are unlikely to be strongly revisionary. Seen in this light, theories of bounded rationality are not, at least for the most part, normative replacements for the Standard Picture.[1] Their primary contribution is to provide new answers to procedural questions about rational inquiry, because it is at the level of inquiry rather than its results that our bounds are most strongly felt.

The RRCV does not, in the first instance, blame agents or theories, but rather suggests a third way that we can react to descriptive deviations from the Standard Picture. This third way involves blaming theorists, not theories, and blaming attitudes, not agents. Let us see in more detail what that entails.

The reason-responsive consequentialist view blames theorists rather than theories because it does not say that the Standard Picture gives a false theory of rational attitudes.[2] Rather, the RRCV reminds us that it is at least as important to ask normative questions at other focal points. In particular, to get a full picture of why agents often deviate from Standard Picture axioms, we need to ask questions about rational inquiry, and these questions are not answered by the Standard Picture. If there is blame to be assigned, it lies not with the Standard Picture for telling us that certain attitudes are irrational, but rather with theorists for failing to follow that judgment with the vindicatory claim that many of these attitudes resulted

---

[1] For other ways of squaring bounded rationality with the Standard Picture, see Berg (2014) and Sturm (2019).

[2] To say that the RRCV does not directly force advocates of the Standard Picture to revise their view of rational belief is not to deny that other phenomena might put pressure on the Standard Picture as applied to belief. For example, you might think that it is not within agents' abilities to believe all tautologies. When combined with (Rational OIC), this would imply that agents cannot be rationally required to believe all tautologies. Here there are many moves to be made, either in defense of the Standard Picture, or in weakening the Standard Picture.

from fully rational inquiry by bounded agents and could only have been avoided through irrational forms of inquiry. It is the job of theorists as much as theories to ensure that the questions we ask show appropriate sensitivity to cognitive bounds.

The RRCV blames attitudes rather than agents because it concedes that Standard Picture violations may involve irrational attitudes, but does not in any obvious way blame agents for those attitudes. We can ask normative questions not only about inquiries and attitudes, but also about agents. Are the agents described in this book rational agents? Do their inquiries spring from cognitive virtue or vice? Can agents be blamed for attitudes that result from rational inquiry? The RRCV does not directly answer any of these questions, but the discussion in this book suggests reasons for answering many of these questions in a positive light. Agents who consistently make the best use of limited cognitive and resources and abilities to achieve their goals may well be rational agents; on at least one account, their dispositions are cognitive virtues (Driver 2001); and their attitudes may well be excused from blame on the basis that they were produced by rational deliberation. While it is true that the RRCV takes agents to sometimes hold irrational attitudes, the RRCV does not in any obvious way blame agents for those attitudes, or conclude that the agents who hold these attitudes are irrational.

In this section, we have seen how the reason-responsive consequentialist view suggests a third way between the options of blaming descriptive deviations on agents or theories. That third way blames not theories, but theorists for focusing too much on normative questions about attitudes rather than processes. The third way also blames attitudes rather than agents, allowing that agents sometimes hold irrational attitudes but blocking any direct inference to the conclusion that the agents themselves are irrational or lack virtue. But the third way is not for everyone. In the next section, we will see that there is at least one important objection to this third way.[3]

---

[3] One interesting question involves defining the job description or theoretical role of rationality on my account. Some theorists distinguish between two non-descriptive projects for rationality: giving guidance and providing normative assessment. While I don't make this distinction myself, a referee proposes that it may be instructive to ask how the distinction between substantive and procedural rationality stacks up to the distinction between normative assessment and guidance-giving. Fixing, for concreteness, the distinction between normative assessment and guidance-giving in a leading recent book (Bermúdez 2009), we see important similarities between the two distinctions. Like Bermúdez's guidance-giving rationality, procedural rationality is closely tied to deliberation; provides first-personal and ability-constrained advice about cognitive processing; and is tightly linked to descriptive theorizing. Like Bermúdez's normative assessment, substantive rationality is all-things-considered, non-subjective, concerned with normative reasons, and may satisfy many traditional requirements of consistency or coherence. There are, also, some divergences between my view of substantive and procedural rationality and Bermúdez's view of guidance-giving and normative assessment: for example, on my view *both* substantive and procedural rationality are all-things-considered, non-subjectivist, and concerned with normative reasons. Does this mean that the distinction between substantive and procedural rationality is a different distinction from what others have meant by the distinction between normative assessment and guidance-giving conceptions of rationality? I am not certain and would leave this question up to the reader to decide.

9.3 The angel's way?

It is often put to me that once we understand just how little impact paradigmatic cognitive bounds have on Standard Picture norms, bounded rationality theorists have no choice but to reject the Standard Picture as an account of rational attitudes. There are at least two ways to motivate this suggestion.

First, Chapter 2 introduced the claim that bounds matter by contrasting theories of bounded rationality to an unbounded approach defended by Ralph Wedgwood. On this approach, a good way to think about rationality is to imagine that an agent's limited cognitive faculties have been scooped out and replaced with an angelic advisor:

> One picturesque way of conceiving of [the] rational probability function is to imagine an *angel* perched inside the thinker's head—where the angel's advice to the thinker takes the form of this rational probability function. Unfortunately, this angel is uncertain about many empirical propositions about the world. However, the angel knows all relevant truths about the mental states and events that are present in the thinker's mind at the time; and she can assign probabilities to these empirical propositions by relying on what she knows about these mental states and events, together with everything that the essential nature of these mental states and events either guarantees or makes likely to be true.
>
> (Wedgwood 2018, p. 99)

Chapter 2 reminded us that humans are not angels, but creatures of flesh and blood. We have the faculties that we have, and rationality requires us to make the best use of these faculties, not the faculties of an angel.

However, if we leave the Standard Picture in place as an account of rational belief, then we will not have landed far from the picture of an angelic advisor. The Standard Picture requires both angels and humans to form arbitrarily complex beliefs and carry out arbitrarily complex deductions. On an information-sensitive view, allowance is made for agents' limited information, but that allowance is also present in Wedgwood's tale of an empirically ignorant angel. These similarities between Wedgwood's account and my account of rational attitudes may be grounds for concern. Insofar as theories of bounded rationality are meant to take us away from stories about angels, we might have expected the apple to land further from the tree.

Another concern is that on my view, normative questions about attitudes are important and well-formed. It makes sense to ask how you ought to deliberate about buying coffee, but also to ask whether you should intend to buy coffee. Then to say that the Standard Picture may be nearly correct about rational attitudes is to cede an important piece of normative ground to unbounded theorists. Insofar as humans are bounded agents, it might be objected that all of the most

important normative questions about cognition should make clear the importance of cognitive bounds, whereas on my view those bounds are primarily felt on processes rather than attitudes.

Where might these objections lead us? The suggestion, I take it, is to adopt a fully or partially indirect approach on which bounds on the cognitive processes that produced an attitude are taken to be relevant to the rational status of the attitude itself. For example, if a belief was formed by rational heuristic deliberation and could only have been avoided by angelic calculations, then we should take that belief to be rational rather than irrational in order to fully reflect the normative importance of cognitive bounds. We saw in Chapter 2 that Simon himself took this route, holding that beliefs are rational when they result from rational deliberation.

The objection is not without force, and readers sympathetic to indirect approaches are welcome to read this book with those approaches in mind. But let the buyer beware. In the next two sections, I argue that combining indirect normative assessment with serious attention to cognitive bounds will have two surprising normative consequences: that most traditional attitudinal norms are false (Section 9.4), and that many attitudinal norms are non-fundamental (Section 9.5). In each case, we will see, these consequences are not speculative or hypothetical: they are the actual, avowed consequences of our best indirect approaches to bounded theorizing. Readers willing to stomach these results are welcome to take an indirect approach. But first I would like to remind the buyer of the consequences of that approach.

9.4 Falsifying traditional attitudinal norms

The first worry for indirection is that it threatens to falsify most traditional norms governing attitudes. The basic worry is that as we have seen, rational processes of inquiry by bounded agents can on occasion produce attitudes which come arbitrarily far apart from the dictates of traditional attitudinal norms. In these cases, indirect normative theories generate pressure against even weakened forms of traditional attitudinal norms, because they generate cases in which rational processes can produce attitudes violating even weakened attitudinal norms.

Let us first see how this problem arises in epistemology (Thorstad 2021). Jane Friedman has recently argued that plausible norms of inquiry may falsify most traditional epistemic norms governing belief (Friedman 2020). For example, Friedman considers weak versions of evidentialism and knowledge norms governing belief:

(EP$_a$) If one has excellent evidence for p at t, then one is permitted to judge p at t.

(KP$_a$) If one is in a position to come to know p at t, then one is permitted to come to know p at t.

Friedman points out that even if we have excellent evidence for *p* or are in a position to come to know *p*, it may still be irrational to engage in the processes of inquiry required to form the belief that *p*. Consider, for example:

> **(Chrysler)** Charlie the contractor has been hired to replace the windows on the Chrysler Building in Manhattan. As a first step, Charlie sits down outside the Chrysler Building to count the number of windows on the building.

While Charlie is counting windows, there are many propositions that Charlie is in a position to know, or for which he has excellent evidence. For example, Charlie is in a position to know that there are fourteen pigeons outside the main entrance to the Chrysler building. But it would be irrational for Charlie to stop counting windows and start counting pigeons, because it is more important for Charlie to count windows than pigeons.

Now on a direct normative theory, this verdict is perfectly compatible with evidentialist or knowledge-based norms governing belief (Thorstad manuscript b). Although it would be irrational for Charlie to engage in the inquiries needed to form the belief that there are fourteen pigeons outside the Chrysler building, the result of this wastefully irrational inquiry might well be a fully rational belief. But Friedman reads EP_a and KP_a as constraints not on the attitudes that we can hold, but rather on the processes of inquiry we can permissibly use to form them. With this interpretation in mind, Friedman quite rightly concludes that EP_a and KP_a are false. It would be irrational to engage in wasteful inquiries such as counting pigeons, even though those inquiries are sanctioned by EP_a and KP_a.

At this point, it might be objected that genuine epistemic norms are not blanket permissions to form any attitude meeting certain conditions, but rather conditions that must be met by attitudes that we do go on to form. So for example, we might defend the converses of EP_a and KP_a:

> **(EP$_a'$)** If one is permitted to judge *p* at *t*, then one has excellent evidence for *p* at *t*.
>
> **(KP$_a'$)** If one is permitted to judge *p* at *t*, then one is in a position to come to know *p* at *t*.[4]

It is no objection to EP_a' that Charlie has excellent evidence for the belief that there are fourteen pigeons outside the Chrysler building, or to KP_a' that Charlie is in a position to know this. After all, EP_a' and KP_a' do not say that it would be rational to form any belief which Charlie has excellent evidence for, or is in a position to

---

[4] Of course, KP_a' is not a strict converse of KP_a, because KP_a deals with what we are permitted to come to know rather than judge. But the strict converse of KP_a is not the principle we are after; that principle would be almost trivially true.

know; only that any belief it would be rational for Charlie to form meets these conditions.

But indirect theories of bounded rationality cause trouble for these converse norms as well. Suppose you are presented with two German cities and asked to judge which of the two cities is the largest. You might make that judgment using a heuristic called *take the best* (TTB). TTB instructs agents to order items of information by their validity as predictors of city size (Gigerenzer 1991; Gigerenzer and Goldstein 1996).[5] Perhaps the most important information is the designation as a national capital, followed by other cues such as being the site of a major exposition or having a major-league soccer team. TTB instructs agents to move downwards through cues in order of predictive validity, until they find a cue which tells in favor of one city being larger. For example, although neither Hamburg nor Essen is a national capitol, Hamburg, unlike Essen, has a major league soccer team. So TTB instructs agents to judge that Hamburg is larger than Essen, and in fact Hamburg is several times larger than Essen.

In many circumstances, heuristics such as TTB may be rational ways to form comparative judgments (Gigerenzer and Goldstein 1996; Hogarth and Karelaia 2006; Martignon and Hoffrage 2002). But on an indirect approach, the rationality of heuristics such as TTB causes trouble for EP'_a and KP'_a. Imagine that some city X outperforms Essen on a single highly ranked item of information, such as the presence of a soccer team, but radically underperforms Essen on most lower-ranked cues. Then it may well be the case that you have excellent evidence for the claim that Essen is larger than X, and are in a position to know that Essen is larger than X by considering all of your information together. But TTB instructs you to halt deliberation quickly with the opposite judgment: that X is larger than Essen.

As before, on a direct approach this instruction is not in conflict with EP'_a and KP'_a. To say that it is rational to inquire using TTB is not to say that if TTB produces an evidentially unsupported belief, this belief itself is rational. But on an indirect approach, EP'_a and KP'_a constrain processes of inquiry: it could not be rational to inquire in a way that produces evidentially unsupported beliefs, or beliefs we are not in a position to know. These verdicts do come into direct tension with the rationality of TTB, insofar as rational use of TTB may sometimes produce beliefs without the properties demanded by EP'_a and KP'_a. Again, we see that even weak forms of evidentialism, knowledge norms, and other traditional epistemic norms will be false.

We might hold out hope that indirect approaches to bounded rationality are less revisionary in the practical sphere than in the epistemic sphere. But that is not the

---

[5] The predictive validity of a cue is the probability that it discriminates correctly, given that it discriminates at all. More formally, fix an environment S of cities and a cue C. For a pair (s_1, s_2) of cities, write $C(s_1 > s_2)$ when C favors s_1 being larger than s_2, and $s_1 > s_2$ when s_1 is in fact larger than s_2. Then the validity of C over S is $|\{(s_1, s_2) \in S \times S : C(s_1 > s_2) \wedge s_1 > s_2\}|$ / $|\{(s_1, s_2) \in S \times S : C(s_1 > s_2)\}|$.

case. As in the epistemic case, it can be rational to engage in processes of inquiry whose results sometimes violate traditional norms governing attitudes, and it can be wastefully irrational to engage in processes of practical inquiry whose resulting attitudes would be sanctioned as rational by traditional norms. Both results cause trouble for indirect approaches, which aim to bring together the rational statuses of inquiries and the attitudes that they produce.

The most familiar cases are cases in which rational processes of practical inquiry produce irrational attitudes. Recall the loving baker from Chapter 3, who holds back five loaves of bread every day for her family because she is moved directly by her deep love for her family. Most normative theories hold that the baker's intention to save five loaves of bread is irrational, for example because it would have better consequences for the baker's family if she were to sell more bread or because the baker has a duty to take reasonable opportunities to provide for her family. But many normative theories also hold that the baker's inquiry may be fully rational. By allowing herself to be directly moved by familial love in deciding how to act, the baker helps her family to feel loved and incorporated in her business concerns, ensures that her family's welfare will be salient in important deliberations, and helps to motivate herself to go out and sell bread every morning. If that is right, then here we have a case of rational inquiry leading to irrational attitudes, a result which indirection threatens to block.

In the other direction, it is widely accepted that some attitudes may be perfectly rational, but could only result from wastefully irrational inquiry. Consider the lifeguard from Chapter 8 who must decide which of several drowning swimmers to save first. Many theories will hold that the lifeguard has a duty to intend to save some particular swimmer first, for example because she is most likely to drown. But we saw in Chapter 8 that the best way for the lifeguard to deliberate may be through a simple heuristic, such as intending to save the nearest swimmer first. The lifeguard may be rationally required to deliberate in this way, in order to keep deliberation time to a minimum. But as a result, she may intend to save a different swimmer than the swimmer she is rationally required to save, or intend to save. Here we have a case of boundedly rational inquiry leading to irrational attitudes, exactly the sort of result that an indirect approach was introduced to block.

So far, we have seen that an indirect approach to the study of bounded rationality threatens to falsify traditional epistemic norms such as evidentialism and knowledge norms, as well as traditional practical norms such as consequentialist and deontological approaches to rational intention. This result may not deter the indirect theorist. After all, if we set out to construct a theory that lands comfortably far from the Standard Picture, then we should not be surprised to learn that this theory comes apart from other traditional normative views as well. But there is a second threat which may trouble even many indirect theorists: on an indirect approach, many attitudinal norms are normatively non-fundamental.

9.5   The fundamentality of attitudinal norms

A second consequence of indirect approaches is that they make attitudinal norms derivative, rather than fundamental in status. As an explanatory matter, indirect views make normative facts about processes explanatorily prior to normative facts about attitudes. And in normative theorizing, they make it difficult to say anything precise about the attitudes agents are required to hold without simply asking how agents are rationally required to inquire, then checking which attitudes would result from those inquiries.

Consider a standard indirect approach, such as Simon's substitute interpretation of procedural rationality (Simon 1976). On this view, the attitudes that result from inquiry are rational just in case they were produced by a rational process of inquiry. As we have seen, the outputs of rational inquiry are quite sensitive to features of agents and their environment. This means that on an indirect approach, it is hard to say anything precise about the rationality of attitudes without first assessing the rationality of the processes that produced them. And while indirect views are not, strictly speaking, committed to any claims about the explanatory priority of norms governing attitudes and processes, these views make it very natural to say that attitudes are rational because they result from rational processes, and that the rationality of processes of inquiry grounds and explains the rationality of the resulting attitudes. It is very hard to see what else could ground or explain attitudinal norms in a way that would guarantee the truth of indirect normative theories, and I do not know of any indirect theorists who have denied that attitudinal norms are grounded and explained by more fundamental procedural norms.

Even weaker forms of indirection often struggle to make room for fundamental attitudinal norms. As we modify accounts of rational attitudes to make them sensitive to bounds governing processes of inquiry, we begin to feel increasing pressure to take attitudinal norms as derivative on more fundamental procedural norms.

By way of illustration, consider the phenomenon of rational delay: rational agents take time to update their beliefs based on new evidence. It is clear that rational delay should be built into accounts of rational inquiry, since processes of inquiry are activities that take time to execute. It might seem natural to build rational delay into accounts of rational belief, so that beliefs may rationally fail to reflect new evidence during an interval of rational delay. Abelard Podgorski (2017) recently investigated the consequences of building rational delay into accounts of rational belief. Podgorski argues convincingly that building rational delay into accounts of rational belief should lead us to regard all attitudinal norms as normatively non-fundamental.

Traditional accounts of rational belief require beliefs to instantly reflect newly acquired evidence. For example, we might hold:

(**Conditionalization**) An agent is rationally required, for any time t_1 and t_2, where $t_1 < t_2$, to have credence at t_2 in P equal to her conditional credence at t_1 in $P|E$, where E is the total evidence she acquires from t_1 to t_2.

(Podgorski 2017, p. 10)

Conditionalization makes no allowance for rational delay. If an agent acquires evidence just milliseconds before t_2, then at t_2 she is rationally required to have fully incorporated that evidence into her beliefs.

We might try to weaken conditionalization to build in an interval of rational delay, during which it is acceptable for agents to fail to update on new evidence. For example, we might defend:

(**Better Conditionalization**) An agent is rationally required, if she has newly acquired evidence E at t_1, to have credence P at t_2 (some appropriate time after t_1) equal to her conditional credence at t_1 in $P|E$.   (Podgorski 2017, p. 10)

But the apparent fundamentality of requirements like Better Conditionalization erodes when we ask a specific question: precisely how long is the interval of rational delay?

It would not do to pick a constant interval of delay for all tasks. Factors like the cognitive difficulty of updating on a given piece of information, the importance of being right about p, and the number and importance of competing tasks should plausibly affect the interval of rational delay. But for just this reason, it is hard to see what else we could say about rational delay without simply reading the interval of delay off from the runtime of rational processes of inquiry. For example, we could try to derive the interval of rational delay from the amount of newly acquired evidence E, the complexity of E or the complexity of the proposition P. But each of these accounts leaves out many of the factors just used to reject a constant interval of rational delay.

And now it looks like the most helpful thing we can do is to first ask which cognitive processes an agent is rationally required to engage in to update her beliefs, taking into account factors such as the cost of those processes, the stakes, and the time required to execute processes. Then we will set the interval of rational delay to the precise amount of time required to execute the process of updating which agents are rationally required to engage in. And then it seems that Better Conditionalization is not a fundamental rational norm in its own right. The fundamental norms of rationality tell agents which processes of updating they are required to use. Those processes span a length of time, from t_1 to t_2, and produce a credence $cr(P|E)$. There is no need to posit a new fundamental norm of Better Conditionalization instructing agents to update from adopt $cr(P|E)$ by t_2, because it already follows from more fundamental procedural norms that they will do so.

Now the phenomenon of rational delay is just one of many bounds on inquiry affecting the beliefs that a rational inquirer will form. As we build these bounds into other attitudinal norms beyond Better Conditionalization, these other norms will begin to look explanatorily non-fundamental in just the same way that Better Conditionalization does. If procedural norms already specify the processes of inquiry that it is rational for agents to use, and if we can read off features of attitudinal norms from these underlying procedural norms, then parsimony seems to demand that we take the procedural norms as fundamental and treat attitudinal norms as derivative on these more procedural norms.

> As long as we have norms about when [a] process is triggered, when it can be interrupted, and how it concludes . . . it follows as a mere corollary that a rational agent who is in the state that triggers the process will form an attitude that is the output of the process at a time . . . So the process view [on which fundamental epistemic norms govern processes] has every indication of being explanatorily fundamental in relation to state-oriented diachronic norms.
>
> (Podgorski 2017, p. 17)

On this view, attitudinal norms may well express genuine normative requirements. But those requirements are not fundamental norms in their own right. Rather, they are the scars or shadows left by more fundamental procedural norms:

> In the same way our physical actions produce scars on our bodies, our mental activities produce attitudes in our minds. But scars, and correspondingly attitudes themselves, are not . . . norm-governed—only the activities that give rise to and manage them are. If fully rational agents, in virtue of their rationality, manifest some regularity in their attitudes, it is because those patterns are the shadows cast by properly functioning processes.   (Podgorski 2017, p. 14)

In this way, building bounds on inquiry such as rational delay into attitudinal norms pushes us toward the same conclusion reached by traditional indirect accounts: that attitudinal norms are derivative rather than fundamental.

When we see the extent to which indirect approaches to bounded rationality make attitudinal norms non-fundamental, some of the motivation for going indirect begins to erode. Our initial concern was that direct approaches make room for a separate and important class of attitudinal norms that do not deeply reflect an agent's bounds. The worry was that theories of bounded rationality should say how agents' bounds bear on all of the most important aspects of their cognition. But the indirect approach does not set out to capture a separate and important class of attitudinal norms that reflect an agent's bounds. The indirect approach makes attitudinal norms less distinct and fundamental, seeking to assimilate these norms

to more fundamental norms governing processes. In the process, much of the distinctness, importance, and interest of attitudinal norms has been lost.

9.6 Conclusion

In this chapter, we saw how the reason-responsive consequentialist view charts a third way between blaming agents and theories for descriptive violations of Standard Picture norms (Section 9.2). My account blames attitudes, not agents for irrationality. It also blames theorists, not theories for emphasizing the irrationality of attitudes instead of the rationality of the processes that produced them. This raised the objection that my approach is too concessive to the Standard Picture because it leaves the Standard Picture largely intact as a theory of rational attitudes (Section 9.3). We could avoid these consequences by adopting an indirect normative theory.

However, I argued that my concessive approach is needed to avoid two normative consequences of indirect normative theories: that most traditional attitudinal norms are false (Section 9.4), and that all attitudinal norms are normatively nonfundamental (Section 9.5). We saw that there is substantial pressure for indirect normative theorists to accept both of these conclusions. And we used case studies from the work of Jane Friedman and Abelard Podgorski to show how indirect views have been used to argue for both conclusions. Readers who are willing to stomach both of these consequences are welcome to read this book through the lights of indirect normative theory. But for my part, I find making peace with the Standard Picture a small price to pay for avoiding both of these consequences.

This chapter discussed the relationship between bounded rationality and the Standard Picture.[6] The last and final chapter takes up the relationship between bounded rationality and recent work in the epistemology of inquiry. Jane Friedman has argued that epistemology must take a *zetetic turn* from the study of doxastic attitudes toward the study of inquiry (Friedman 2020, forthcoming). The zetetic turn in epistemology looks very much like a special case of Simon's procedural turn from the study of attitudes toward the study of inquiry. I argue

---

[6] There are other differences between bounded rationality and the Standard Picture worth exploring. Some may have passed by without notice: unlike the Standard Picture, theories of bounded rationality take a heuristic, process-focused, and ecological approach to rationality (Chapters 2, 8). Some differences are dispositional: bounded rationality theorists tend to hold out more hope for vindicatory epistemology than many advocates of the Standard Picture do (Chapters 2, 7). Other differences run very deep: for example, unlike the Standard Picture, theories of bounded rationality are not architecturally neutral (Chapter 1) and often resist axiomatic characterization or find axiomatization less helpful (Gigerenzer 2019). Some remaining issues, such as the sense in which bounded rationality is a non-maximizing or non-optimizing approach (Elster 1983; Klein 2001; Russell and Wefald 1991) have not been discussed here and would be fruitful avenues for future research.

that we should indeed regard the zetetic turn in epistemology as a special case of the procedural turn. I use this view to draw lessons for the epistemology of inquiry by applying what we have learned about bounded rationality. I also argue that treating the zetetic turn in epistemology as part of a broader procedural turn reveals the need for a second zetetic turn within practical philosophy. I conclude with some guidance for the shape of this second zetetic turn.

10

The zetetic turns

10.1 The zetetic turn and the procedural turn

Epistemology has taken a *zetetic turn* (Friedman 2020, forthcoming). Traditionally, epistemologists have followed a doxastic paradigm concerned primarily with doxastic attitudes:

> According to the doxastic paradigm, epistemic norms are norms that bear almost exclusively on having, forming, revising, maintaining (etc.) beliefs and other belief-like attitudes.   (Friedman forthcoming)

Taking the zetetic turn means shifting much of our attention away from attitudes and toward the processes of inquiry that produce them:

> Taking the zetetic turn . . . means moving to thinking of the norms of epistemology as speaking to the entire process of inquiry.   (Friedman forthcoming)

If I have done my job right, these words will sound welcome and familiar.

In Chapter 2, we saw that Herbert Simon took the fundamental turn in the study of bounded rationality to be the *procedural turn* from substantive to procedural rationality. Chapters 2, 8, and 9 defended a view on which substantive rationality asks questions about the rationality of attitudes, whereas procedural rationality asks questions about the processes of inquiry that produce them. If this is right, then the zetetic turn looks very much like a special case of the procedural turn.[1] Taking the zetetic turn means moving from the study of doxastic attitudes toward the study of theoretical inquiry, whereas taking the procedural turn means moving from the study of doxastic and practical attitudes toward the study of theoretical and practical inquiry. My aim in this chapter is to take this claim seriously and see what follows.

Section 10.2 argues that recent work in the epistemology of inquiry, or *zetetic epistemology* should be treated as an episode in the study of bounded rationality. If this is right, then it will make sense to treat the zetetic turn in epistemology as part

---

[1] And if this is wrong, then the zetetic turn may still look like a special case of the procedural turn. We will just have to reinterpret both turns in a consistent way, for example through the lights of indirect normative assessment.

of a broader procedural turn within the study of bounded rationality. Section 10.3 uses this insight to draw lessons for zetetic epistemology by applying what we have learned about bounded rationality.

Thinking about the zetetic turn as part of a broader procedural turn reminds us that there is not one zetetic turn, but two: a turn in epistemology from doxastic attitudes toward theoretical inquiry, and a second turn in practical philosophy from practical attitudes toward practical inquiry. Section 10.4 argues that taking the procedural turn seriously requires taking a second zetetic turn in practical philosophy, then uses what we have learned about bounded rationality to draw lessons for the zetetic turn in practical philosophy. Section 10.5 concludes.[2]

10.2 Zetetic epistemology and bounded rationality

What is the relationship between zetetic epistemology and bounded rationality? In this book, I have argued that all human rationality is bounded rationality. We have minds and bodies with specific capacities. These capacities limit the inquiries that are possible for us and impose costs on inquiry. Our minds and bodies are located in environments which interact with our cognitive and physical architecture to further restrict the inquiries we can engage in, and to structure the costs and benefits of these inquiries. Chapters 2 and 8 defended a relevance constraint on which paradigmatic bounds bear on how humans are rationally required to cognize.

If all human rationality is bounded rationality, then there can be no question of whether zetetic epistemology is a type of bounded rationality. While it may sometimes make methodological sense to highlight certain bounds over others in our analysis, in general zetetic epistemology must be conceived as a type of bounded rationality if it is to be a theory of human rationality. Readers sympathetic to this view may skip ahead to the next section, which draws lessons for zetetic epistemology from what we have learned about bounded rationality.

But some readers may think that in addition to bounded rationality, there is a distinct type of ideal rationality that abstracts away from bounds such as minds, bodies, or environments (Carr 2022; Richter 1990; Smithies 2015). Granting for the sake of argument that this program gets a take on some subjects, perhaps such as rational belief, we might ask whether it makes sense to develop a theory of ideally rational inquiry. Here, I think that whatever our views in the case of belief and other attitudes, it does not make good sense to think of recent work in zetetic

---

[2] A brief caution before I begin: the material in this chapter will be of most interest to readers with an existing interest in the epistemology of inquiry. Readers with more general interests may benefit less from some of this discussion.

epistemology as concerned with a phenomenon that can be captured by theories of ideal rationality.

Here is the problem in brief. Many of the most central and paradigmatic challenges that arise for inquirers arise in a nontrivial way because we are bounded. These challenges would either fail to arise, or else become nearly unrecognizable if the bounds generating them were removed. If this is right, then it does not make sense to think of recent work in zetetic epistemology, which has been concerned with specific types of challenges and aimed to deliver recognizable classes of solutions, as concerned with a type of ideal rationality that abstracts away from cognitive bounds.

To see the point, consider four challenges at issue in recent zetetic epistemology (Table 10.1). The first is evidence-gathering (Hall and Johnson 1998; Smith 2014; Woodard and Flores forthcoming). When are agents rationally required to gather evidence during inquiry, and what evidence should they gather? This challenge arises because we are informationally bounded: there is relevant information that we do not have. Because we lack relevant information, it may make sense to gather that information.

Next, consider double-checking (Christensen 2007; Friedman 2019a; Hawthorne and Stanley 2008; Woodard forthcoming). Recent work has stressed that rational agents should sometimes double-check their beliefs in order to ensure that they have not made a mistake. When and how should agents double-check their beliefs? This question arises because we are fallible. We need to double-check our beliefs because we could have made a mistake the first time.

Turn now to junk belief (Friedman 2018; Harman 1986; Michaelian 2011). Recent work has stressed that agents often have duties to avoid forming, storing, or retaining 'junk' or trivial beliefs. Obligations to avoid junk belief arise because of two bounds. First, inquiry is costly: forming junk beliefs consumes valuable resources such as time and effort. Second, memory is bounded: we may be in danger of running out of storage space in memory (Harman 1986), and we are certainly in danger of cluttering memory retrieval, with the result that future inquiries will be guided by a high proportion of irrelevant information. On this basis, it has been argued not only that we should avoid forming junk beliefs or storing them in long-term memory, but also that we may be rationally required

Table 10.1 Challenges for inquirers and their motivating bounds

Challenge for inquirers	Motivating bound(s)
Gathering evidence	Informational bounds
Double-checking	Fallibility
Junk belief	Memorial bounds, costs of inquiry
Allocating attention	Attentional bounds, computational bounds

to forget junk beliefs in order to declutter our long-term memories (Hertwig and Engel 2021; Schooler and Hertwig 2005).

Finally, inquirers must choose how to allocate attention between competing internal and external stimuli (Siegel 2017; Sims 2003). Paying attention to the right stimuli ensures that our inquiries are guided by the most relevant features of our minds and environments. This challenge arises because we are attentionally bounded: we cannot attend to all stimuli at once. It may also arise because we are computationally bounded, unable to process all of the information that an attentionally unbounded agent would have before them.

To see that these challenges need to be conceived as challenges facing bounded agents, let us ask what would happen to these challenges if the generating bounds were removed. Here the answer is that each challenge would either fail to arise, or arise in a strange and unrecognizable way.

Turn first to evidence-gathering. When are informationally unbounded agents rationally required to gather evidence, and what evidence should they gather? Here there is no challenge at all. Informationally unbounded agents should never gather evidence, and in fact they could not do so. They already have all evidence at their disposal.

Now consider double-checking. When and how should infallible agents double-check their beliefs? While it is possible for infallible agents to double-check their beliefs, they should never do so. An agent who could not have been wrong the first time has no need to double-check her beliefs.

The story is much the same with junk belief. Consider an agent with unbounded memory capacities for storage and retrieval, and for whom inquiry is not costly. What would be wrong with such an agent forming, storing, or failing to forget junk beliefs? Plausibly, there would be nothing wrong with this.[3] If forming and storing junk beliefs is costless, storage space is unlimited, and junk beliefs will not clutter later memory retrieval, then it is hard to see what could be the matter with junk belief. In fact, such an agent would plausibly be rationally required to form and store all junk beliefs with a nonzero probability of proving useful during her lifetime.

Finally, consider the allocation of attention. Which stimuli should an attentionally unbounded agent attend to? Here the answer is that she should attend to all stimuli at once, at least if she is also unbounded in her capacity to process incoming information. An unbounded agent may benefit from attending to each additional stimulus, and can never be harmed in the process.

The lesson here is that even if we take some features of agency to be governed by separate types of bounded and ideal rationality, we should not think of norms

---

[3] Michaelian (2011) demurs, holding that the presence of junk beliefs makes an agent's belief state intrinsically worse. But Michaelian holds this in large part because he is unwilling to accept the bounded rationality explanation of duties to avoid junk belief. If we take that explanation on board, then there is much less explanatory pressure to posit a new primitive type of badness attaching to junk belief.

of inquiry, in the sense raised by recent work in zetetic epistemology, as tracking a type of ideal rationality. Many of the most central and paradigmatic challenges facing rational inquirers only arise in an interesting way because we are bounded, and would fail to arise or else become nearly unrecognizable when the relevant bounds were removed. In this sense, it makes good sense to think of zetetic epistemology as studying a type of bounded rationality, and to apply what we have learned about bounded rationality to draw lessons for zetetic epistemology. I take up this task in the next section.

10.3 Lessons for zetetic epistemology

Thinking about the zetetic turn in epistemology as part of a broader procedural turn within the study of bounded rationality allows us to apply what we have learned about bounded rationality to shed light on zetetic epistemology. In this section, I show how thinking about the procedural turn yields new motivations for taking the zetetic turn in epistemology (Section 10.3.1); raises new questions within zetetic epistemology (Section 10.3.2); highlights the relevance of novel normative considerations (Section 10.3.3); yields an important methodological lesson (Section 10.3.4); and provides a new view of rational inquiry (Section 10.3.5).

10.3.1 Motivating the zetetic turn

We saw in Chapter 4 that recent work in zetetic epistemology has given three motivations for the zetetic turn. First, it is argued, a wide range of historical work in epistemology has been concerned with inquiry (Friedman 2017b; Mattherne manuscript; Misak 1987; Striker 2001). In taking the zetetic turn, we pick up a project that history has conceived of as central to the epistemological enterprise: understanding the nature and norms of inquiry.

Second, our attitudes at any given time reflect ongoing processes of temporally extended inquiry. If we focus only on attitudes held at moments in time, rather than the processes of inquiry that produce and modify these attitudes, we risk developing a temporally parochial epistemology of the present moment that is unable to capture the ongoing processes of inquiry through which attitudes are shaped (Friedman 2020).

Finally, on some views the norms governing theoretical inquiry are norms of theoretical rationality (Kelly 2003; Sylvan manuscript; Thorstad 2021). It is often thought that theoretical rationality falls within the domain of epistemology, rather than practical philosophy. Hence if it is correct to say that norms of theoretical inquiry are norms of theoretical rationality, then developing these norms is a task for epistemologists.

Viewing the zetetic turn in epistemology as part of a broader procedural turn reveals a powerful new motivation for taking the zetetic turn: the zetetic turn is an essential component of a humanly adequate epistemology. We saw in Chapter 8 that theories of bounded rationality should be process-focused because many paradigmatic cognitive bounds are felt most strongly as bounds on inquiry, rather than as bounds on the attitudes that result from inquiry. And we saw in Chapter 9 that if we focus only on norms governing attitudes, we will recover a picture much closer to the Standard Picture than to existing theories of bounded rationality. Insofar as we want to develop a humanly adequate epistemology, we need to place more emphasis on norms governing inquiry in order to get a full and complete picture of how bounds bear on rational cognition. As part of a broader procedural turn, the zetetic turn in epistemology is a fundamental and indispensable component of the study of bounded rationality.

10.3.2 New questions for zetetic epistemology

In addition to providing new motivations for the zetetic turn, existing work on bounded rationality also raises novel questions for zetetic epistemologists. There is a rich body of scholarly questions about rational inquiry in the literature on bounded rationality, and many of these questions have yet to be taken up by zetetic epistemologists. Any of these questions would be ripe for further exploration. I give two examples below.

First, we saw in Chapters 2 and 8 that bounded rationality is process-focused, aimed primarily at studying the processes of inquiry used by bounded agents. We saw also that many of the processes used by bounded inquirers are heuristics. If that is right, then it is important to ask questions about the nature of rational heuristic inquiry in order to build a specific and detailed account of rational inquiry.

For example, we might ask which heuristics are available to bounded inquirers (Gigerenzer and Gaissmaier 2011; Gigerenzer and Selten 2001). When and why is it rational to use a given heuristic strategy (Lieder and Griffiths 2017; Marewski and Schooler 2011)? In which ways do rational heuristics differ from the non-heuristic strategies studied by many epistemologists? And what changes must be made to the traditional epistemological narrative in order to accommodate the rationality of heuristic inquiry (Bishop and Trout 2004; Karlan 2021; Weinberg et al. 2001)?

Second, we saw in Chapters 2 and 8 that rationality is ecological, or environment-relative. Because heuristics and other strategies perform well in some environments and poorly in others, we must never say of a strategy that it is rational or irrational full-stop. Instead, we must ask in which environments a strategy would be rational to use. We saw in Chapter 2 that ecological rationality is

a methodological thesis as well as a normative thesis. As theorists, we need to strive as far as possible to build the structure of environments into normative models or we risk going awry. For example, we may erroneously claim that strategies are irrational when they are instead rational responses to environmental constraints that fall outside of our normative models.

One natural way to build the structure of environments into normative models would be to ask how the rationality of particular strategies varies with the structure of the environment. For example, we might ask whether and how the accessibility of evidence in an agent's environment affects her duties to gather evidence. Or we might ask which features of environments help to transform processes of double-checking belief from a salutary safeguard into an irrational pathology. Taking ecological rationality seriously urges us to put such questions at the forefront of our research agenda.

10.3.3 New normative considerations

In the last subsection, we saw that thinking about bounded rationality raises new questions for zetetic epistemology. At the same time as bounded rationality raises new questions for zetetic epistemology, it also highlights a broad range of normative considerations that are relevant to answering these questions.

We saw in Chapter 2 that paradigmatic bounds such as limited cognitive abilities, costs of inquiry, and the structure of the environment bear on the rationality of inquiry. And we saw in Chapter 5 that the noncognitive stakes of inquiry, tradeoffs that arise during inquiry, and the need to avoid rationalizing problematic inferences by stereotyping also constrain accounts of rational inquiry. We also saw in Chapter 5 that not all theories of rational inquiry give plausible accounts of how these factors bear on the rationality of inquiry. This finding suggests two lessons for zetetic epistemology.

First, we must develop views which take normative considerations such as limited abilities, costs, environments, stakes, tradeoffs, and stereotyping seriously. If a view struggles to deliver detailed or plausible suggestions for how inquirers should respond to these normative considerations, then that is reason to deepen or change the view.

Second, these normative considerations are not ancillary to the study of bounded rationality. They capture some of the most fundamental challenges facing bounded inquirers. If that is right, then questions about these normative considerations need to be placed at the forefront of research into zetetic epistemology. For example, we might ask how the costs of inquiry affect the rationality of closing inquiry, or how the noncognitive stakes of inquiry affect the aims that it is rational for inquirers to have. By putting prominent focus on questions like these, we will become more likely to develop theories of rational

inquiry that can give detailed and accurate accounts of how bounded agents should respond to the normative considerations which structure their lives.

10.3.4 Methodological implications

We saw in Chapters 2 and 8 that many claims about bounded rationality are not only normative claims, but also methodological claims. They remind us as theorists that it is important to model paradigmatic cognitive bounds in as much detail as possible in order to ensure that the normative impact of these bounds is correctly reflected in our theories.

Above, we discussed the methodological importance of building models which reflect the structure of the environment. But humans are not only limited by our environments. We are also limited by our internal cognitive structure, which restricts the processes that are within our power and imposes costs on many of those processes. Theories of bounded rationality stress that it is important to engage in detail with empirical psychology in order to build a detailed and accurate story of rational human inquiry. If we do not do this, we risk building models which have little to do with the actual situation of human inquirers. We will then make normative mistakes when we project these models onto human agents, wrongly classifying those agents as irrational when they are instead responding rationally to cognitive constraints that lie outside of our normative models.

The methodological importance of attending to descriptive psychology is perhaps best revealed in a letter from Herbert Simon to the economist Ariel Rubinstein (Simon 1998). Rubinstein is well-known for constructing a dazzling array of sophisticated mathematical models, each exploring the consequences of relaxing a different modeling assumption from neoclassical economics. For a small sampling, Rubinstein gives us models of agents who treat similar options as identical; agents who play coordination games with near-complete common knowledge; agents whose attitudes are distributed across non-partitional information structures; agents who absentmindedly forget information; and agents who choose which information to retain about consumer prices (Rubinstein 1989, 1998, 2012).

The problem, according to Simon, is that Rubinstein's models make only tangential reference to the actual structure of human cognition:

> Aside from the use that you make of the Tversky-Kahneman experiments . . . almost the only reference to empirical matters that I detect in your pages is an occasional statemen[t] like 'a casual observation' and 'the phenomenon exhibited here is quite common.'   (Simon 1998, p. 188)

While it is quite right to use empirical research to motivate the program of bounded rationality modeling, Simon holds, this itself is not enough. We cannot simply allude to the need for bounded rationality modeling, then write down any model which strikes our fancy. We need to use empirical findings to help us to see which modeling assumptions are the right ones to make, and which are wrong.

> At the moment we don't need more models; we need evidence that will tell us what models are worth building and testing... [Models which describe] not all of the phenomena that we can imagine, but those that actually occur.
>
> (Simon 1998, p. 190)

The right approach, for Simon, is to use descriptive psychology to deliver detailed and accurate characterizations of the actual structure of human cognition, then use these findings to build models which correctly describe the situation facing human cognizers and the normatively correct response to that situation.

The lesson for zetetic epistemology is much the same. We can set out and study any number of challenges facing bounded inquirers. But if we do not describe these challenges in a way that incorporates detailed and accurate characterizations of human psychology, then we will end up developing models that are either too abstract to capture the situation of bounded inquirers or else are specific enough to make false assumptions about the situation facing human inquirers. Then we will be at a high risk of drawing false normative conclusions by projecting models onto human cognizers which do not capture relevant cognitive bounds.

10.3.5 A new view of rational inquiry

One final way in which thinking about bounded rationality can contribute to zetetic epistemology is by introducing a new view of rational inquiry: the reason-responsive consequentialist view. On that view, acting rationally consists in doing what we have most reason to do in response to the reasons for which we ought to do it. What we have most reason to do is to promote value, and the promotion of value is assessed in an information-sensitive way. Since inquiry is an activity, an account of rational inquiry falls out of the RRCV as a special case. The RRCV incorporates further elements such as a sharp level-separation between questions about belief and inquiry (Thorstad 2021); a rich axiology on which many achievements, both intellectual and non-intellectual, bear final value; and a consequentializing program aimed at delivering plausible and non-revisionary explanations of a range of normative phenomena.

Chapter 5 suggested that a good way to test the reason-responsive consequentialist view is by thinking about what the RRCV can say about specific cases of rational inquiry. In Chapter 5, we saw what the RRCV has to say about stakes, tradeoffs, and stereotypes. In Chapter 6, we saw what the RRCV says about clutter avoidance, norms of friendship, and norms of logical omniscience. In Chapter 7, we saw what the RRCV has to say about anchoring and adjustment, as well as some experiments testing conditional reasoning. In Chapters 8–9, we saw what the RRCV has to say about the normative importance of cognitive bounds, procedural rationality, the rationality of heuristic inquiry, and ecological rationality.

These case studies leave many stones unturned. Recent work in zetetic epistemology has raised many interesting questions about rational inquiry which have not been discussed in detail in this book. When is it rational for agents to gather evidence (Hall and Johnson 1998; Smith 2014; Woodard and Flores forthcoming)? When is it rational for agents to double-check their beliefs once those beliefs are formed (Christensen 2007; Friedman 2019a)? Is there a sense in which agents are rationally required to suspend judgment during inquiry, and if so what is that sense (Friedman 2017b; McGrath 2021)? It would be an interesting project, going forward, to see what the RRCV says about these questions. If the RRCV continues to deliver plausible, explanatorily powerful and intuitively non-revisionary verdicts in these cases, these results will at once shed light on specific cases and provide some evidence for the RRCV. On the other hand, if the RRCV struggles in these cases, then that will be some evidence against the RRCV, and this evidence might be used to construct a new theory.

In this section, we have seen how treating the zetetic turn in epistemology as a special case of the procedural turn in bounded rationality yields five lessons for zetetic epistemology. Applying what we have learned about bounded rationality introduces a new motivation for taking the zetetic turn: taking the zetetic turn is essential to the construction of a humanly adequate epistemology. Thinking about bounded rationality also reveals new questions for zetetic epistemology, such as questions about heuristic inquiry and the structure of the environment. Theories of bounded rationality highlight a range of relevant normative considerations that theories of rational inquiry should be sensitive to, including cognitive abilities, costs, environments, stakes, tradeoffs, and stereotyping. Thinking about bounded rationality highlights the methodological importance of attention to descriptive psychology, and introduces a new view of rational inquiry: the reason-responsive consequentialist view.

So far, we have been concerned with the zetetic turn in epistemology. But the zetetic turn in epistemology is a special case of Simon's more general procedural turn. To complete the zetetic turn, we need to take an analogous turn within practical philosophy. Section 10.4 characterizes this second zetetic turn, showing how the characteristic claims about bounded rationality made in Chapter 2 carry over into the practical domain, and how the five lessons we have just learned generalize from theoretical to practical inquiry.

10.4 The other zetetic turn

10.4.1 Two zetetic turns

Inquiries come in two types. On the one hand, there are theoretical inquiries. Theoretical inquiries terminate, at least often, in the formation of doxastic attitudes such as belief. And perhaps theoretical inquiries originate with the formation of a special type of interrogative attitude such as curiosity or wondering (Friedman 2013; Newton manuscript; Whitcomb 2010). On the other hand, there are practical inquiries. Practical inquiries terminate, at least often, in the formation of practical attitudes such as plans, intentions, and preferences. Perhaps practical inquiries begin with the formation of an intention, such as the intention to decide which of a set of options to take, although I do not mean to commit to any particular view on the matter.

The procedural turn is a turn from the study of attitudes toward the study of the processes of inquiry that produce them. One half of the procedural turn is the zetetic turn in epistemology from doxastic attitudes to theoretical inquiry. But our practical attitudes such as plans and intentions are no less the product of inquiry than our doxastic attitudes are. For this reason, taking the procedural turn also involves a second zetetic turn in practical philosophy, from practical attitudes to practical inquiry. Insofar as we are committed to studying bounded rationality, we need to think seriously about practical inquiry.

If we want to get a handle on the rationality of practical inquiry for bounded agents, where might we begin? While there are many places we might begin, one place to start would be with the observation that this book is not a book about theoretical inquiry. It is a book about inquiry, both practical and theoretical. Although some of the discussion is couched in terms familiar to epistemologists, most of the main lessons carry over to the practical case without change. For this reason, we can make a good start at understanding the rationality of practical inquiry by applying what we have already learned about inquiry in this book.

By way of illustration, in the rest of this section I show how the five characteristic claims about bounded rationality developed in Chapters 2, 7, and 8 can be defended as claims about practical inquiry (Section 10.4.2). Then I show how the same five lessons drawn in Section 10.3 for zetetic epistemology can be carried over to the case of practical inquiry (Section 10.4.3).

10.4.2 Five characteristic claims

Chapter 2 introduced five characteristic claims about bounded rationality: that bounds matter; rationality is process-focused, ecological and often heuristic; and many challenged areas of human cognition can be vindicated as in an important

sense rational. Below, I show how each of these claims can be defended when we are concerned with practical rather than theoretical inquiry.

Begin with the claim that bounds matter: paradigmatic cognitive bounds such as limited cognitive abilities and the costs of computation bear on the rationality of agents' judgment and decisionmaking. This is already a claim about decision-making, or practical inquiry, and rightly so. Practical inquiries, like theoretical inquiries, are limited by our abilities and incur costs. In practical philosophy, it is widely accepted that ought implies can, and insofar as this is true we cannot be required to inquire in ways that are beyond our abilities. Likewise, in practical philosophy it is generally accepted that the noncognitive costs of actions bear on the rationality of those activities. Insofar as inquiry is an activity, this means that the noncognitive costs of inquiry bear on its rationality. And once we accept the relevance of noncognitive costs, it is hard to find a reason to deny the normative relevance of cognitive costs.

Turn next to the claim that theories of rationality should be process-focused. The argument given in Chapter 8 for this claim was that standard theories of rational attitudes have very little to do with paradigmatic cognitive bounds. I illustrated this argument by considering the case of an agent choosing among a long shelf full of artisanal vinegars at the supermarket. I suggested that it would be rational for the agent to make up her mind about which vinegar to buy using a frugal heuristic, even though that heuristic would likely lead her to intend to buy a bottle of vinegar that is not expectedly best. In this case, we would have a case of rational inquiry leading to irrational intention, but the intention would only be irrational because relevant bounds such as the costs of cognition are entirely irrelevant to the rationality of the agent's intention. Once we see how little paradigmatic cognitive bounds have to do with the rationality of practical attitudes such as intention, we should accept that procedural facts often give a more helpful and complete picture of the normative lives of bounded agents.

Note well that this argument is already phrased in terms of practical inquiry. While we could construct analogous cases to illustrate the importance of a process-focused epistemology, the argument given in Chapter 8 is already an argument for adopting a process-focused practical philosophy.

The third characteristic claim was that rational agents often deliberate using a toolbox of fast-and-frugal heuristic strategies (Gigerenzer and Gaissmaier 2011). This claim can be defended on all of the same grounds as the rationality of heuristic belief formation. First, the accuracy-effort tradeoff in theoretical inquiry becomes a quality-effort tradeoff in practical inquiry. In many cases, the quality of decisions trades off against the effort of making them, and heuristic processes often strike the best balance between quality and effort (Johnson and Payne 1985; Thorstad 2021). Second, our cognitive abilities limit the practical inquiries that we can engage in, just like they limit the theoretical inquiries we can engage in. Sometimes the only inquiries we can engage in are heuristic, and in such cases

we cannot be rationally required to inquire nonheuristically because we cannot do so. Finally, in practical inquiry less can also be more: in some cases, frugal heuristics reliably outperform more complex procedures, even ignoring the cost of decisionmaking, because they are less prone to overgeneralize on sparse data (Geman et al. 1992; Gigerenzer and Brighton 2009). In such cases, the rationality of heuristic inquiry should be widely conceded.

Fourth, we held that rationality is ecological, or environment-relative. Just as the structure of task environments affects features such as the speed, cost, and accuracy of theoretical inquiries, so too the structure of the environment affects the speed, cost and decision quality of practical inquiries. For this reason, we must never say of a process of practical inquiry that it is rational or irrational full-stop. We must always ask: in which environments would this process be rational to use?

Finally, Chapter 2 made the vindicatory claim that large swaths of seemingly irrational experimental findings can be vindicated as the results of rational inquiry. Although the discussion of vindicatory epistemology in Chapter 7 focused on theoretical inquiry, we have already met one well-known example of vindicatory practical philosophy. Chapter 5 reviewed a range of findings from the psychology of poverty and took those findings to suggest that it is often rational for the poor to engage in short-termist patterns of cognitive resource allocation, devoting increased resources such as attention, and computational bandwidth toward pressing short-term problems and consequences. We saw that the predictable consequence of rational short-termist resource allocation will be that agents sometimes intend to, and do overborrow, undersave, or fail to comply with medical instructions. Chapter 5 suggested that these behaviors and the intentions that produced them may be irrational, but that these intentions and behaviors are often the results of fully rational practical inquiry, and could only be avoided by irrational forms of practical inquiry. If that is right, then the psychology of poverty provides a good case study for how a vindicatory practical philosophy could be developed.

In this subsection, we have seen how the five characteristic claims about bounded rationality developed in Chapter 2 can be advanced as claims about practical inquiry as well as claims about theoretical inquiry. In the rest of this section, I show how the five lessons for zetetic epistemology drawn out in Section 10.3 can be recast as guidance for normative theories of practical inquiry.

10.4.3 Five lessons

Section 10.3 used what we have learned about bounded rationality to draw five lessons for the study of theoretical inquiry. Below, I show how each of these lessons can be recast as lessons about practical inquiry.

Motivations

In Section 10.3, we saw that the study of bounded rationality generates a powerful new motivation for taking the zetetic turn in epistemology. The zetetic turn in epistemology is a special case of the more general procedural turn. And the procedural turn is the fundamental turn in the study of bounded rationality. For that reason, taking the zetetic turn in epistemology is essential to the construction of a humanly adequate epistemology, one that is concerned with bounded agents rather than inhuman angels. But the zetetic turn in practical philosophy is also a special case of the procedural turn. For this reason, taking the zetetic turn in practical philosophy is just as essential to the construction of a humanly adequate practical philosophy as the zetetic turn in epistemology is essential to the construction of a humanly adequate epistemology.

Questions

In Section 10.3, we saw that the study of bounded rationality generates new questions for zetetic epistemology. By way of example, we surveyed two sets of questions about heuristics and environments, respectively. All of these questions have practical analogs that deserve careful scrutiny.

In the practical domain, we can ask questions about heuristics: which heuristics are available to bounded agents; when are these heuristics rational; and what changes should be made to practical philosophy to accommodate the rationality of heuristic inquiry? We can also ask questions about environments. For example, how does the availability of information in the environment affect the rationality of gathering information before making a decision? And how does the structure of the environment bear on the boundary between rational and irrational reopenings of previously settled practical inquiries?

Normative considerations

In Section 10.3, we saw that the study of bounded rationality raises at least six normative considerations that theories of rational inquiry should be sensitive to: costs, abilities, environments, stakes, tradeoffs, and stereotyping. These considerations are no less relevant to practical inquiry than they are to theoretical inquiry.

Practical inquiries incur costs, and these costs bear on the rationality of practical inquiry because they are part of its outcomes. Practical inquiries are limited by our abilities, and since ought implies can, we cannot be rationally required to engage in practical inquiries that are beyond our abilities. Environments alter the reward structure of practical inquiries, and hence the rationality of engaging in them.

Practical inquiries should be stakes-sensitive: we need to spend more time thinking about how to pay our taxes than how to pay a dinner tab. Practical inquiries incur tradeoffs, for example by drawing resources away from other practical inquiries, and these tradeoffs affect the rationality of engaging in tradeoff-inducing processes. And practical inquiries can involve objectionable

stereotypes. For example, an executive can hand his coat to a woman in the boardroom because she is a woman. It is important to develop accounts of practical inquiry which say that such stereotype-involving inquiries are often irrational.

Methodology

In Section 10.3, we saw that it is important to build detailed specifications of internal cognitive bounds into models of theoretical inquiry. If we try to make do with less-detailed, or less-informed models, we risk making false claims about the rationality of theoretical inquiry by projecting conclusions from those models onto agents who rationally respond to bounds that lie outside of our normative models. Exactly the same conclusion holds for practical inquiry. Cognitive bounds matter to the rationality of practical inquiry. If we do not theorize on the basis of detailed and precise models of cognitive bounds, we risk making false claims about the rationality of practical inquiry by projecting conclusions from those models onto agents who rationally respond to bounds that lie outside of our normative models.

A new view

In Section 10.3, we saw that the study of bounded rationality contributes a new view of theoretical inquiry: the reason-responsive consequentialist view. But the RRCV is not only an account of theoretical inquiry. It is an account of rational action, of which practical inquiry is a subspecies.

All of the main components of the RRCV are familiar to practical philosophers: a global-consequentialist account of rightness, a reason-responsiveness view of rationality, and an information-sensitive account of deontic modals. In fact, many of these components were developed by practical philosophers. And two of the subsidiary components of the RRCV were introduced by consequentialists: a sharp level separation between inquiry and intention, and a consequentializing program aimed at recovering consequentialist explanations of a range of normative data.

As in the case of theoretical inquiry, what is novel about the reason-responsive consequentialist view is the combination of these components together with the bounded rationality program underlying them. I suggested that a good test of the RRCV would be to see what it says about problem cases for theoretical inquiry, and that is equally true when we turn our attention to practical inquiry. For example, we might ask when it is rational for agents to gather evidence before making decisions; whether practical inquiry involves a type of suspension of judgment; or when it is rational to reopen previously closed practical inquiries. If the RRCV continues to deliver plausible, explanatorily powerful and intuitively non-revisionary verdicts in these cases, these results will at once shed light on specific cases and provide some evidence for the RRCV. On the other hand, if the RRCV struggles in these cases, then that will be some evidence against the RRCV, and this evidence might be used to construct a new theory.

10.5　Conclusion

We have come a long way. It is time to take stock. Part 1 introduced the Standard Picture and five characteristic claims associated with theories of bounded rationality. To ground, clarify and apply those claims, we needed a theory of rational inquiry. Part 2 developed such a view, the reason-responsive consequentialist view, and answered an objection that the RRCV is problematically non-epistemic. Part 3 gave two arguments for the RRCV: the argument from minimal criteria, and the explanatory argument.

The business of Part 4 has been to apply the work done in Parts 1–3 to shed light on bounded rationality, the Standard Picture, and zetetic epistemology. Chapters 7–8 used the RRCV to ground, clarify and apply the claims about bounded rationality developed in Part 2. Chapter 9 revisited the relationship between bounded rationality and the Standard Picture, arguing that bounded rationality and the Standard Picture should for the most part be taken as compatible theories of different objects.

The concern of this last and final chapter has been with the zetetic turn. We saw that epistemology has taken a zetetic turn from the study of doxastic attitudes toward the study of theoretical inquiry (Section 10.1), and that this turn can be helpfully viewed as a special case of the more general procedural turn in bounded rationality (Section 10.2). This generated five lessons for zetetic epistemology (Section 10.3). We also saw that completing the procedural turn requires taking a second zetetic turn in practical philosophy, from the study of practical attitudes toward the study of practical inquiry (Section 10.4). We saw that most of the leading claims of this book can be rephrased as claims about practical inquiry, including the five characteristic claims about bounded rationality from Chapter 2, and the five lessons for zetetic epistemology drawn in Section 10.4.

If we have gone this far, then what is next? The discussion in this chapter suggests two places that we might go from here. First, although practical philosophers have often been concerned with inquiry, practical philosophy does not find itself in the midst of anything like the full-blown zetetic turn that has gripped recent epistemology. If I am right that it is important for normative theorizing to take the procedural turn, and that taking the procedural turn requires taking a second zetetic turn within practical philosophy, then we have excellent reason to begin in earnest to take the zetetic turn in practical philosophy.

Second, we saw in Sections 10.3–10.4 that the study of bounded rationality yields five lessons for the study of rational inquiry. Taking bounded rationality seriously reminds us that the zetetic turns are necessary to the construction of a humanly adequate normative philosophy. The study of bounded rationality introduces new questions, such as questions about heuristics and environments. It also introduces new normative considerations that theories of rational inquiry should be sensitive to. The study of bounded rationality reminds us of the importance

of enriching normative models with detailed information about agents' internal cognitive architecture. And thinking about rationality brings forth a new view, the reason-responsive consequentialist view, which can help us to resolve questions about rational inquiry. Taking these lessons to heart helps us to see what a humanly adequate normative theory could be, and what remains to be done along the way toward humanizing normative theory.

References

Abdelbar, Ashraf M. and Hedetniemi, Sandra M. 1998. "Approximating MAPs for belief networks is NP-hard and other theorems." *Artificial Intelligence* 102:21–38.

Adams, Ernest. 1962. "On rational betting systems." *Archiv für mathematische Logik und Grundlagenforschung* 6:112–28.

Adams, Robert Merrihew. 1976. "Motive utilitarianism." *Journal of Philosophy* 73:467–81.

Alston, William. 1978. "Meta-ethics and meta-epistemology." In Alvin Goldman and Jaegwon Kim (eds), *Values and morals*, 275–97. D. Reidel.

Anderson, Alan Ross and Belnap, Nuel D. 1975. *Entailment: The logic of relevance and necessity*, volume 1. Princeton University Press.

Anderson, John. 1990. *The adaptive character of thought*. Psychology Press.

Anderson, John and Schooler, Lael. 1991. "Reflections of the environment in memory." *Psychological Science* 2:396–408.

Arango-Muñoz, Santiago and Michaelian, Kourken. 2014. "Epistemic feelings and epistemic emotions." *Philosophical Inquiries* 2:97–122.

Arkes, Hal and Blumer, Catherine. 1985. "The psychology of sunk cost." *Organizational Behavior and Human Decision Processes* 35:124–40.

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

Armendt, Brad. 1993. "Dutch books, additivity, and utility theory." *Philosophical Topics* 21:1–20.

Arpaly, Nomy. 2017. "Epistemology and sandwiches." https://theviewfromtheowlsroost. com/2017/10/29/epistemology-and-sandwiches/.

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

Audi, Robert. 1983. "The causal structure of indirect justification." *Journal of Philosophy* 80:398–415.

Audi, Robert. 1986. "Acting for reasons." *Philosophical Review* 95:511–46.

Badhwar, Neera. 1991. "Why it is wrong to be always guided by the best: Consequentialism and friendship." *Ethics* 101:483–504.

Baehr, Jason. 2009. "Evidentialism, vice, and virtue." *Philosophy and Phenomenological Research* 78:545–67.

Bar-Hillel, Maya. 1973. "On the subjective probability of compound events." *Organizational Behavior and Human Performance* 9:396–406.

Bar-Hillel, Maya and Neter, Efrat. 1993. "How alike is it versus how likely is it: A disjunction fallacy in probability judgments." *Journal of Personality and Social Psychology* 65:1119–31.

Bar-Hillel, Maya and Wagenaar, Willem. 1991. "The perception of randomness." *Advances in Applied Mathematics* 12:428–54.

Benartzi, Shlomo and Thaler, Richard. 2007. "Heuristics and biases in retirement savings behavior." *Journal of Economic Perspectives* 21:81–104.

Berg, Nathan. 2014. "The consistency and ecological rationality approaches to normative bounded rationality." *Journal of Economic Methodology* 21:375–95.

Berker, Selim. 2013. "Epistemic teleology and the separateness of propositions." *Philosophical Review* 122:337–93.

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

Bernecker, Sven. 2018. "On the blameworthiness of forgetting." In Kourken Michaelian, Dorothea Debus, and Denis Perrin (eds), *New directions in the philosophy of memory*, 241–58. Routledge.

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.

Betzler, Monika and Schroth, Jörg. 2019. "The good of consequentialized deontology." In Christian Seidel (ed.), *Consequentialism: New directions, new problems*, 115–35. Oxford University Press.

Birnbaum, Michael H. 2008. "New paradoxes of risky decision making." *Psychological Review* 115:463–501.

Bishop, Michael and Trout, J.D. 2004. *Epistemology and the psychology of human judgment*. Oxford University Press.

Bjork, Robert A. and Whitten, William B. 1974. "Recency-sensitive retrieval processes in long-term free recall." *Cognitive Psychology* 6:173–89.

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.

Boghossian, Paul. 2014. "What is inference?" *Philosophical Studies* 169:1–18.

Bolinger, Renée. 2020. "The rational impermissibility of accepting (some) racial generalizations." *Synthese* 197:2415–31.

Bolker, Ethan. 1966. "Functions resembling quotients of measures." *Transactions of the American Mathematical Society* 124:292–312.

BonJour, Lawrence. 1985. *The structure of empirical knowledge*. Harvard University Press.

Boult, Cameron. 2021. "There is a distinctively epistemic kind of blame." *Philosophy and Phenomenological Research* 103:518–34.

Bradley, Richard. 2017. *Decision theory with a human face*. Cambridge University Press.

Brandt, Richard. 1959. *Ethical theory*. Prentice-Hall.

Briggs, R.A. 2009. "Distorted reflection." *Philosophical Review* 118:59–85.

Broome, John. 2013. *Rationality through reasoning*. Wiley.

Brown, Jessica. 2020. "What is epistemic blame?" *Noûs* 54:389–407.

Buchak, Lara. 2013. *Risk and rationality*. Oxford University Press.

Buchak, Lara. 2016. "Decision theory." In Christopher Hitchcock and Alan Hájak (eds), *Oxford Handbook of Probability and Philosophy*. Oxford University Press.

Buehler, Roger, Griffin, Dale, and Ross, Michael. 1994. "Exploring the 'planning fallacy': Why people underestimate their task completion times." *Journal of Personality and Social Psychology* 67:366–81.

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, 178–83.

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.

Carr, Jennifer. 2015. "Subjective ought." *Ergo* 2:678–710.

Carr, Jennifer. 2022. "Why ideal epistemology?" *Mind* 131:1131–62.

Chapman, Gretchen and Johnson, Eric. 1994. "The limits of anchoring." *Journal of Behavioral Decision Making* 7:223–42.

Chapman, Loren and Chapman, Jean. 1959. "Atmosphere effect re-examined." *Journal of Experimental Psychology* 58:220–6.

Charlow, Nate. 2013. "What we know and what to do." *Synthese* 190:2291–323.

Charlow, Nate. 2016. "Decision theory: Yes! Truth conditions: No!" In Nate Charlow and Matthew Chrisman (eds), *Deontic modality*, 47–81. Oxford University Press.

Chase, William and Simon, Herbert. 1973. "Perception in chess." *Cognitive Psychology* 4:55–81.

Chater, Nick and Oaksford, Mike. 1999a. "The probability heuristics model of syllogistic reasoning." *Cognitive Psychology* 258:191–258.

Chater, Nick and Oaksford, Mike. 1999b. "Ten years of the rational analysis of cognition." *Trends in Cognitive Sciences* 3:57–65.

Cheng, Patricia and Holyoak, Keith. 1985. "Pragmatic reasoning schemas." *Cognitive Psychology* 17:391–416.

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

Chickering, David Maxwell, Heckerman, David, and Mack, Christopher. 2004. "Large-sample learning of Bayesian networks is NP-hard." *Journal of Machine Learning Research* 5:1297–1330.

Chisholm, Roderick. 1956. "The concept of empirical evidence." *Journal of Philosophy* 53:722–31.

Chong, Dennis and Druckman, James. 2007. "Framing theory." *Annual Review of Political Science* 10:103–26.

Christensen, David. 1996. "Dutch-book arguments depragmatized: Epistemic consistency for partial believers." *Journal of Philosophy* 93:450–79.

Christensen, David. 2001. "Preference-based arguments for probabilism." *Philosophy of Science* 68:356–76.

Christensen, David. 2007. "Epistemology of disagreement: The good news." *Philosophical Review* 116:187–217.

Christensen, David. 2021. "The ineliminability of epistemic rationality." *Philosophy and Phenomenological Research* 103:501–17.

Clifford, William. 1877. "The ethics of belief." *Contemporary Review* 29:283–309.

Cohen, Jonathan. 1981. "Can human irrationality be experimentally demonstrated?" *Behavioral and Brain Sciences* 4:317–31.

Cohen, Sheldon, Doyle, William, and Baum, Andrew. 2006. "Socioeconomic status is associated with stress hormones." *Psychosomatic Medicine* 68:414–20.

Cohen, Stewart. 2016a. "Replies to my commentators." *Inquiry* 59:906–22.

Cohen, Stewart. 2016b. "Theorizing about the epistemic." *Inquiry* 59:839–57.

Conee, Earl. 2016. "The epistemic." *Inquiry* 59:858–66.

Cooke, Roger. 1991. *Experts in uncertainty: Opinion and subjective probability in science.* Oxford University Press.

Cosmides, Leda. 1989. "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task." *Cognition* 31:187–276.

Cosmides, Leda and Tooby, John. 1996. "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty." *Cognition* 58:1–73.

Dagum, Paul and Luby, Michael. 1993. "Approximating probabilistic inference in Bayesian belief networks is NP-hard." *Artificial Intelligence* 60:141–53.

Daoust, Marc-Kevin. 2021. "Adversariality and ideal argumentation: A second-best perspective." *Topoi* 40:887–98.

Davidson, Donald. 1980. *Essays on actions and events*. Clarendon Press.

Davidson, Donald, McKinsey, John Charles Chenoweth, and Suppes, Patrick. 1955. "Outlines of a formal theory of value, I." *Philosophy of Science* 22:140–60.

Dawes, Robyn and Corrigan, Bernard. 1974. "Linear models in decision making." *Psychological Bulletin* 81:95–106.

De Bona, Glauber and Staffel, Julia. 2017. "Graded incoherence for accuracy-firsters." *Philosophy of Science* 84:189–213.

De Bona, Glauber and Staffel, Julia. 2018. "Why be (approximately) coherent?" *Analysis* 78:405–15.

De Neys, Wim, Rossi, Sandrine, and Houdé, Olivier. 2013. "Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools." *Psychonomic Bulletin and Review* 20:269–73.

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.

Denes-Raj, Veronika and Epstein, Seymour. 1994. "Conflict between intuitive and rational processing: When people behave against their better judgment." *Journal of Personality and Social Psychology* 66:819–29.

DeRose, Keith. 2009. *The case for contextualism*. Oxford University Press.

Desvousges, William, Johnson, Reed, Dunford, Richard, Boyle, Kevin, Hudson, Sara, and Wilson, Nicole. 1992. *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy*. Research Triangle Institute.

Dijksterhuis, Ap and Nordgren, Loran. 2006. "A theory of unconscious thought." *Perspectives on Psychological Science* 1:95–109.

Dotson, Kristie. 2011. "Tracking epistemic violence, tracking practices of silencing." *Hypatia* 26:236–56.

Dotson, Kristie. 2018. "Accumulating epistemic power: A problem with epistemology." *Philosophical Topics* 46:125–54.

Dotson, Kristie. 2019. "Tales from an apostate." *Philosophical Issues* 29:69–83.

Doughtery, Trent. 2012. "Reducing responsibility: An evidentialist account of epistemic blame." *European Journal of Philosophy* 20:534–47.

Doughtery, Trent. 2014. "The ethics of belief is ethics (period): Reassigning responsibilism." In Jonathan Matheson and Rico Vitz (eds), *The ethics of belief*, 146–68. Oxford University Press.

Dreier, Jamie. 2011. "In defense of consequentializing." *Oxford Studies in Normative Ethics* 97–119.

Driver, Julia. 2001. *Uneasy Virtue*. Cambridge University Press.

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

Dulany, Don and Hilton, Denis. 1991. "Conversational implicature, conscious representation, and the conjunction fallacy." *Social Cognition* 9:85–110.

Dunn, Michael and Restall, Greg. 2002. "Relevance logic." In Dov Gabbay and F. Guenther (eds), *Handbook of philosophical logic*, volume 6, 1–128. Springer, 2nd edition.

Eddy, David. 1982. "Probabilistic reasoning in clinical medicine: Problems and opportunities." In Daniel Kahneman, Paul Slovic, and Amos Tversky (eds), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Einhorn, Hillel and Hogarth, Robin. 1975. "Unit weighting schemes for decision making." *Organizational Behavior and Human Performance* 13:171–92.

Elqayam, Shira. 2018. "The new paradigm in psychology of reasoning." In Linden Ball and Valerie Thompson (eds), *Routledge international handbook of thinking and reasoning*, 130–50. Routledge.

Elqayam, Shira and Over, David. 2013. "New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over." *Thinking and Reasoning* 19:249–65.

Elster, Jon. 1983. *Sour grapes: Studies in the subversion of rationality*. Cambridge University Press.

Enqvist, Sebastian. 2012. "Interrogative belief revision based on epistemic strategies." *Studia Logica* 100:453–79.

Epley, Nicholas and Gilovich, Thomas. 2001. "Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors." *Psychological Science* 12:391–6.

Epley, Nicholas and Gilovich, Thomas. 2004. "Are adjustments insufficient?" *Personality and Social Psychology Bulletin* 30:447–60.

Epley, Nicholas and Gilovich, Thomas. 2006. "The anchoring-and-adjustment heuristic: Why the adjustments are insufficient." *Psychological Science* 17:311–8.

Eva, Benjamin and Hartmann, Stephan. 2018. "Bayesian argumentation and the value of logical validity." *Psychological Review* 125:806–21.

Evans, Gary and Schamberg, Michelle. 2009. "Childhood poverty, chronic stress, and adult working memory." *Proceedings of the National Academy of Sciences* 106:6545–9.

Evans, Jonathan, Barston, Julie, and Pollard, Paul. 1983. "On the conflict between logic and belief in syllogistic reasoning." *Memory and Cognition* 11:295–306.

Evans, Jonathan, Handley, Simon, and Over, David. 2003. "Conditionals and conditional probability." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:321–35.

Evans, Jonathan and Lynch, J.S. 1973. "Matching bias in the selection task." *British Journal of Psychology* 64:391–7.

Evans, Jonathan and Stanovich, Keith. 2013. "Dual-process theories of higher cognition: Advancing the debate." *Perspectives on Psychological Science* 8:223–41.

Falbo, Arianna. forthcoming. "Inquiring minds want to improve." *Australasian Journal of Philosophy*.

Falbo, Arianna. ms. "Should epistemology take the zetetic turn?" Manuscript.

Fantl, Jeremy and McGrath, Matthew. 2009. *Knowledge in an uncertain world*. Oxford University Press.

Feldman, Richard. 2000. "The ethics of belief." *Philosophy and Phenomenological Research* 60:667–95.

Feldman, Richard. 2002. "Epistemological duties." In Paul Moser (ed.), *The Oxford Handbook of epistemology*, 362–83. Oxford University Press.

Fine, Kit. 2020. "Yablo on subject matter." *Philosophical Studies* 177:129–71.

Firth, Roderick. 1956. "Ultimate evidence." *Journal of Philosophy* 53:732–9.

Firth, Roderick. 1959. "Chisholm and the ethics of belief." *Philosophical Review* 68:493–506.

Fischhoff, Baruch, Slovic, Paul, Lichtenstein, Sarah, Read, Stephen, and Combs, Barbara. 1978. "How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits." *Policy Sciences* 9:127–52.

Fishburn, Peter. 1974. "Lexicographic orders, utilities and decision rules: A survey." *Management Science* 20:1442–71.

Fleisher, Will. 2022. "Pursuit and inquisitive reasons." *Studies in History and Philosophy of Science* 94:17–30.

Foley, Richard. 1987. *The theory of epistemic rationality*. Harvard University Press.

Foot, Philippa. 1983. "Utilitarianism and the virtues." *Proceedings of the American Philosophical Association* 57:275–83.

Frederick, Shane and Mochon, Daniel. 2012. "A scale distortion theory of anchoring." *Journal of Experimental Psychology: General* 141:124–33.

Fredrickson, Barbara L and Kahneman, Daniel. 1993. "Duration neglect in retrospective evaluations of affective episodes." *Journal of Personality and Social Psychology* 65:45–55.

Frege, Gottlob. 1893/1903. *Grundgesetze der Arithmetik, Band I/II*. Verlag Hermann Pohle.

Fricker, Miranda. 2007. *Epistemic injustice*. Oxford University Press.

Friedman, Jane. 2013. "Question-directed attitudes." *Philosophical Perspectives* 27:145–74.

Friedman, Jane. 2017a. "Inquiry and belief." *Noûs* 53:296–315.

Friedman, Jane. 2017b. "Why suspend judging?" *Noûs* 51:302–326.

Friedman, Jane. 2018. "Junk beliefs and interest-driven epistemology." *Philosophy and Phenomenological Research* 97:568–83.

Friedman, Jane. 2019a. "Checking again." *Philosophical Issues* 29:84–96.

Friedman, Jane. 2019b. "Teleological epistemology." *Philosophical Studies* 176:673–91.

Friedman, Jane. 2020. "The epistemic and the zetetic." *Philosophical Review* 129:501–36.

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

Friedman, Jane. ms. "The aim of inquiry?" Manuscript.

Friedman, Milton. 1953. *Essays in positive economics*. University of Chicago Press.

Furnham, Adrian and Chu Boo, Hua. 2011. "A literature review of the anchoring effect." *Journal of Socio-Economics* 40:35–42.

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* 92:325–36.

Galesic, Mirta, García-Retamero, Rocío, and Gigerenzer, Gerd. 2009. "Using icon arrays to communicate medical risks: Overcoming low numeracy." *Health Psychology* 28:210–6.

García-Retamero, Rocío and Hoffrage, Ulrich. 2013. "Visual representation of statistical information improves diagnostic inferences in doctors and their patients." *Social Science and Medicine* 83:27–33.

Geman, Stuart, Bienenstock, Elie, and Doursat, René. 1992. "Neural networks and the bias/variance dilemma." *Neural Computation* 4:1–58.

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

Gentzen, Gerhard. 1935. "Untersuchungen über das logische Schließen. I." *Mathematische Zeitschrift* 39:176–210.

George, Jennifer and Dane, Erik. 2016. "Affect, emotion, and decision making." *Organizational Behavior and Human Decision Processes* 136:47–55.

Gershman, Samuel, Horvitz, Eric, and Tenenbaum, Joshua. 2015. "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines." *Science* 349:273–8.

Gibbard, Allan. 2003. *Thinking how to live*. Harvard University Press.

Gibbard, Allan and Harper, William. 1978. "Counterfactuals and two kinds of expected utility." In Cliff Hooker, John Leach, and Edward McClennen (eds), *Foundations and applications of decision theory*, 125–62. Reidel.

Gigerenzer, Gerd. 1991. "How to make cognitive illusions disappear: Beyond 'heuristics and biases.'" *European Review of Social Psychology* 2:83–115.

Gigerenzer, Gerd. 1996. "On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1986)." *Psychological Review* 103:592–6.

Gigerenzer, Gerd. 2019. "Axiomatic rationality and ecological rationality." *Synthese* 194:3547–64.

Gigerenzer, Gerd and Brighton, Henry. 2009. "Homo heuristicus: Why biased minds make better inferences." *Topics in Cognitive Science* 1:107–43.

Gigerenzer, Gerd and Gaissmaier, Wolfgang. 2011. "Heuristic decision making." *Annual Review of Psychology* 62:451–82.

Gigerenzer, Gerd and Goldstein, Daniel. 1996. "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review* 103:650–9.

Gigerenzer, Gerd and Hoffrage, Ulrich. 1995. "How to improve Bayesian reasoning without instruction: Frequency formats." *Psychological Review* 102:684–704.

Gigerenzer, Gerd and Selten, Reinhard (eds). 2001. *Bounded rationality: The adaptive toolbox*. MIT press.

Gigerenzer, Gerd and Selten, Reinhard. 2001b. "Rethinking rationality." In Gerd Gigerenzer and Reinhard Selten (eds), *Bounded rationality: The adaptive toolbox*, 1–12. MIT Press.

Gilbert, Daniel and Hixton, Gregory. 1991. "The trouble of thinking: Activation and application of stereotypic beliefs." *Journal of Personality and Social Psychology* 60:509–17.

Gilovich, Thomas and Griffin, Dale. 2002. "Heuristics and biases: Then and now." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds), *Heuristics and biases: The psychology of intuitive judgment*, 1–18. Cambridge University Press.

Gilovich, Thomas, Vallone, Robert, and Tversky, Amos. 1985. "The hot hand in basketball: on the misperception of random sequences." *Cognitive Psychology* 17:325–31.

Gneiting, Tilmann and Raftery, Adrian. 2007. "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102:359–78.

Godden, Duncan and Baddeley, Alan. 1975. "Context-dependent memory in two natural environments: On land and underwater." *British Journal of Psychology* 66:325–31.

Gödel, Kurt. 1929. *Über die Vollständigkeit des Logikkalkuls*. Ph thesis, University of Vienna.

Gödel, Kurt. 1930. "Die vollständigkeit der axiome des logischen funktionenkalküls." *Monatshefte für Mathematik und Physik* 37:349–60.

Goldberg, Sanford. 2016. "On the epistemic significance of evidence you should have had." *Episteme* 13:449–70.

Goldman, Alvin. 1986. *Epistemology and cognition*. Harvard University Press.

Goldstein, Daniel and Gigerenzer, Gerd. 1999. "The recognition heuristic: How ignorance makes us smart." In Gerd Gigerenzer, Peter Todd, and The ABC Research Group (eds), *Simple heuristics that make us smart*, 37–58. Oxford University Press.

González, Richard and Wu, George. 1999. "On the shape of the probability weighting function." *Cognitive Psychology* 38:129–66.

Greaves, Hilary. 2013. "Epistemic decision theory." *Mind* 122:915–52.

Greaves, Hilary and Wallace, David. 2006. "Justifying conditionalization: Conditionalization maximizes expected epistemic utility." *Mind* 115:607–32.

Greco, John. 2011. "Evidentialism about knowledge." In Trent Doughtery (ed.), *Evidentialism and its discontents*, 167–78. Oxford University Press.

Griffiths, Thomas, Chater, Nick, Norris, Dennis, and Pouget, Alexandre. 2012. "How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012)." *Psychological Bulletin* 138:415–22.

Hahn, Ulrike. 2020. "Argument quality in real world argumentation." *Trends in Cognitive Sciences* 24:363–75.

Hall, Richard and Johnson, Charles. 1998. "The epistemic duty to seek more evidence." *American Philosophical Quarterly* 35:129–39.

Halpern, Joseph. 2003. *Reasoning about uncertainty*. MIT Press.

Hampton, Jean. 1994. "The failure of expected-utility theory as a theory of reason." *Economics and Philosophy* 10:195–242.

Harman, Gilbert. 1973. *Thought*. Princeton University Press.

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

Harman, Gilbert. 2004. "Practical aspects of theoretical reasoning." In Alfred R. Mele and Piers Rawling (eds), *The Oxford handbook of rationality*, 45–56. Oxford University Press.

Harsanyi, John. 1977. "Rule utilitarianism and decision theory." *Erkenntnis* 11:25–53.

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

Hattori, Msasi. 2002. "A quantitative model of optimal data selection in Wason's selection task." *Quarterly Journal of Experimental Psychology* 55:1241–72.

Hawthorne, John and Stanley, Jason. 2008. "Knowledge and action." *Journal of Philosophy* 105:571–90.

Hedden, Brian. 2015. "Time-slice rationality." *Mind* 124:449–91.

Heitz, Richard. 2014. "The speed-accuracy tradeoff: history, physiology, methodology, and behavior." *Frontiers in Neuroscience* 8:1–19.

Hertwig, Ralph and Engel, Christoph (eds). 2021. *Deliberate ignorance*. MIT Press.

Hertwig, Ralph and Gigerenzer, Gerd. 1999. "The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors." *Journal of Behavioral Decision Making* 12:275–305.

Hintikka, Jaakko. 2001. "Is logic the key to all good reasoning?" *Argumentation* 15:35–57.

Hoffrage, Ulrich, Lindsey, Samuel, Hertwig, Ralph, and Gigerenzer, Gerd. 2000. "Communicating statistical information." *Science* 290:2261–2.

Hogarth, Robin and Karelaia, Natalia. 2006. "'Take-the-best' and other simple strategies: Why and when they work 'well' with binary cues." *Theory and Decision* 61:205–49.

Houthakker, Hendrik. 1950. "Revealed preference and the utility function." *Economica* 17:159–74.

Howard, Ronald. 1966. "Information value theory." *IEEE Transactions on Systems Science and Cybernetics* 2:22–6.

Howard, Ronald. 1968. "The foundations of decision analysis." *IEEE Transactions on Systems Science and Cybernetics* 4:211–9.

Howes, Andrew, Lewis, Richard, and Vera, Alonso. 2009. "Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action." *Psychological Review* 116:717–51.

Hurley, Paul. 2013. "Consequentializing and deontologizing: Clogging the conseqeuntialist vacuum." *Oxford Studies in Normative Ethics* 3:123–52.

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

Icard, Thomas. 2021. "Why be random?" *Mind* 130:111–39.

Jackson, Frank. 1991. "Decision-theoretic consequentialism and the nearest and dearest objection." *Ethics* 101:461–82.

Jacowitz, Karen and Kahneman, Daniel. 1995. "Measures of anchoring in estimation tasks." *Personality and Social Psychology Bulletin* 21:1161–6.

Jeffrey, Richard. 1965. *The logic of decision*. University of Chicago Press.

Jeffrey, Richard C. 1978. "Axiomatizing the logic of decision." In Cliff Hooker, John Leach, and Edward McClennen (eds), *Foundations and applications of decision theory*, 227–31. Springer.

Johnson, Eric, Hershey, John, Meszaros, Jacqueline, and Kunreuther, Howard. 1993. "Framing, probability distortions, and insurance decisions." *Journal of Risk and Uncertainty* 7:35–51.

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

Johnson, Eric and Schkade, David. 1989. "Bias in utility assessments: Further evidence and explanations." *Management Science* 35:406–24.

Johnson, Joseph and Raab, Markus. 2003. "Take the first: Option-generation and resulting choices." *Organizational Behavior and Human Decision Processes* 91:215–29.

Joyce, James. 1998. "A nonpragmatic vindication of probabilism." *Philosophy of Science* 65:575–603.

Joyce, James. 1999. *The foundations of causal decision theory*. Cambridge University Press.

Joyce, James M. 2009. "Accuracy and coherence: Prospects for an alethic epistemology of partial belief." In Franz Huber and Christoph Schmidt-Petri (eds), *Degrees of belief*, 263–97. Springer.

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.

Kahneman, Daniel. 1981. "Who shall be the arbiter of our intuitions?" *Behavioral and Brain Sciences* 4:339–40.

Kahneman, Daniel. 2003. "Maps of bounded rationality: Psychology for behavioral economics." *American Economic Review* 93:1449–75.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, Daniel, Knetsch, Jack, and Thaler, Richard. 1990. "Experimental tests of the endowment effect and the Coase theorem." *Journal of Political Economy* 98:1325–48.

Kahneman, Daniel and Tversky, Amos. 1973. "On the psychology of prediction." *Psychological Review* 80:237–51.

Kahneman, Daniel and Tversky, Amos. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica* 47:263–92.

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* 39:1249–55.

Karlan, Brett. 2021. "Reasoning with heuristics." *Ratio* 34:100–8.

Karni, Edi and Vierø, Marie-Louise. 2013. " 'Reverse Bayesianism': A choice-based theory of growing awareness." *American Economic Review* 103:2790–2810.

Kawall, Jason. 2013. "Friendship and epistemic norms." *Philosophical Studies* 165:349–70.

Keller, Simon. 2004. "Friendship and belief." *Philosophical Papers* 33:329–51.

Kelly, Thomas. 2002. "The rationality of belief and some other propositional attitudes." *Philosophical Studies* 110:163–96.

Kelly, Thomas. 2003. "Epistemic rationality as instrumental rationality: A critique." *Philosophy and Phenomenological Research* 66:612–40.

Kelp, Christoph. 2021a. *Inquiry, knowledge, and understanding*. Oxford University Press.

Kelp, Christoph. 2021b. "Theory of inquiry." *Philosophy and Phenomenological Research* 103:359–84.

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* 113:12868–73.

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

Kirby, Kris. 1994. "Probabilities and utilities of fictional outcomes in Wason's four-card selection task." *Cognition* 51:1–28.

Klayman, Joshua and Ha, Young-Won. 1987. "Confirmation, disconfirmation, and information in hypothesis testing." *Psychological Review* 94:211–28.

Klein, Gary. 2001. "The fiction of optimization." In Gerd Gigerenzer and Reinhard Selten (eds), *Bounded rationality: The adaptive toolbox*, 103–21. MIT Press.

Knuth, Donald and Moore, Ronald. 1953. "An analysis of alpha-beta pruning." *Artificial Intelligence* 6:293–326.

Kolodny, Niko. 2005. "Why be rational?" *Mind* 114:509–63.

Kolodny, Niko and MacFarlane, John. 2010. "Ifs and oughts." *Journal of Philosophy* 107:115–43.

Koriat, Asher. 1997. "Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning." *Journal of Experimental Psychology: General* 126:349–70.

Koriat, Asher and Ackerman, Rakefet. 2010. "Metacognition and mindreading: Judgments of learning for self and other during self-paced study." *Consciousness and Cognition* 19:251–64.

Koriat, Asher, Ma'ayan, Hilit, and Nussinson, Ravit. 2006. "The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior." *Journal of Experimental Psychology: General* 135:36–69.

Korsgaard, Christine. 1997. "The normativity of instrumental reason." In Garrett Cullity and Berys Gaut (eds), *Ethics and practical reason*, 215–54. Oxford University Press.

Korsgaard, Christine. 2005. "Acting for a reason." *Danish Yearbook of Philosophy* 40:11–35.

Kratzer, Angelika. 1981. "The notional category of modality." In Hans-Jürgen Erikmeyer and Hannes Rieser (eds), *Words, worlds, and contexts*, 38–74. De Gruyter.

Kwisthout, Johan, Wareham, Todd, and van Rooij, Iris. 2011. "Bayesian intractability is not an ailment that approximation can cure." *Cognitive Science* 35:779–84.

Lack, David. 1943. *The life of the robin*. Witherby.

Lavin, Douglas. 2004. "Practical reason and the possibility of error." *Ethics* 114:424–57.

Lee, Carole and Schunn, Christian. 2011. "Social biases and solutions for procedural objectivity." *Hypatia* 26:352–73.

Lee, Wooram. forthcoming. "Belief and settledness." *Inquiry* forthcoming.

Leech, Jessica. 2015. "Logic and the Laws of Thought." *Philosophers' Imprint* 15:1–27.

Leite, Adam. 2007. "Epistemic instrumentalism and reasons for belief: A reply to Tom Kelly's 'Epistemic rationality as instrumental rationality: A critique.'" *Philosophy and Phenomenological Research* 75:456–64.

Leitgeb, Hannes and Pettigrew, Richard. 2010a. "An objective justification of Bayesianism I: Measuring inaccuracy." *Philosophy of Science* 77:201–35.

Leitgeb, Hannes and Pettigrew, Richard. 2010b. "An objective justification of Bayesianism II: The consequences of minimizing inaccuracy." *Philosophy of Science* 77:236–72.

Lerner, Jennifer and Keltner, Dacher. 2001. "Fear, anger, and risk." *Journal of Personality and Social Psychology* 81:146–59.

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

Lever, Annabelle. 2007. "What's wrong with racial profiling? Another look at the problem." *Criminal Justice Ethics* 26:20–8.

Lewis, David. 1999. *Papers in metaphysics and epistemology*, volume 2. Cambridge University Press.

Lieder, Falk and Griffiths, Thomas. 2017. "Strategy selection as rational metareasoning." *Psychological Review* 124:762–94.

Lieder, Falk and Griffiths, Thomas. 2020. "Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources." *Behavioral and Brain Sciences* 43:E1.

Lieder, Falk, Griffiths, Thomas, Huys, Quentin, and Goodman, Noah. 2018. "The anchoring bias reflects rational use of cognitive resources." *Psychonomic Bulletin and Review* 25:322–49.

Linde, Lena and Bergström, Monica. 1992. "The effect of one night without sleep on problem-solving and immediate recall." *Psychological Research* 54:127–37.

Lindsay, D Stephen and Johnson, Marcia K. 1989. "The eyewitness suggestibility effect and memory for source." *Memory and Cognition* 17:349–58.

Lipsey, Richard and Lancaster, Kelvin. 1956. "The general theory of second best." *Review of Economic Studies* 24:11–32.

Littlejohn, Clayton. 2012. *Justification and the truth connection*. Cambridge University Press.

Longino, Helen. 1990. *Science as social knowledge*. Princeton University Press.

Lopes, Lola. 1982. "Toward a procedural theory of judgment." Office of Naval Research Final Report.

Lord, Errol. 2017. "What you're rationally required to do and what you ought to do (are the same thing!)." *Mind* 126:1109–54.

Lord, Errol. 2018. *The importance of being rational*. Oxford University Press.

Lord, Errol. 2020. "Suspension of judgment, rationality's competition, and the reach of the epistemic." In Sebastian Schmidt and Gerhard Ernst (eds), *The ethics of belief and beyond*, 126–45. Routledge.

Lyons, Jack. 2016. "What we talk about when we talk about epistemic justification." *Inquiry* 59:867–88.

MacFarlane, John. 2002. "Frege, Kant, and the logic in logicism." *Philosophical Review* 111:25–65.

MacFarlane, John. manuscript. "Is logic normative for thought?" Manuscript.

Macnamara, John. 1986. *A border dispute*. MIT Press.

Maguire, Barry and Woods, Jack. 2020. "The game of belief." *Philosophical Review* 129:211–49.

Mahtani, Anna. 2012. "Diachronic Dutch book arguments." *Philosophical Review* 121:443–50.

Mahtani, Anna. 2015. "Dutch books, coherence, and logical consistency." *Noûs* 49:522–37.

Mani, Anandi, Mullainathan, Sendhil, Shafir, Eldar, and Zhao, Jiaying. 2013. "Poverty impedes cognitive function." *Science* 341:976–80.

Marcucilli, Paul. 2019. *The appraisal of belief*. PhD thesis, Harvard University.

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

Martignon, Laura and Hoffrage, Ulrich. 2002. "Fast, frugal and fit: Simple heuristics for paired comparison." *Theory and Decision* 52:29–71.

Martignon, Laura and Laskey, Kathryn Blackmond. 1999. "Bayesian benchmarks for fast and frugal heuristics." In Gerd Gigerenzer, Peter Todd, and The ABC Research Group (eds), *Simple heuristics that make us smart*, 169–88. Oxford University Press.

Masny, Michal. 2020. "Friedman on suspended judgment." *Synthese* 197:5009–26.

Mattherne, Matthew. manuscript. "The attitudes of Socratic inquiry." Manuscript.

McCain, Kevin. 2012. "The interventionist account of causation and the basing relation." *Philosophical Studies* 159:357–82.

McCormick, Miriam. 2015. *Believing against the evidence: Agency and the ethics of belief.* Routledge.

McGrath, Matthew. 2016. "Cohen on 'epistemic.'" *Inquiry* 59:889–905.

McGrath, Matthew. 2021. "Being neutral: Suspension of judgement, agnosticism and inquiry." *Noûs* 55:463–84.

McHugh, Conor and Way, Jonathan. 2016. "Fittingness first." *Ethics* 126:575–606.

McKay, Ryan and Dennett, Daniel. 2009. "The evolution of misbelief." *Behavioral and Brain Sciences* 32:493–510.

Meacham, Christopher and Weisberg, Jonathan. 2011. "Representation theorems and the foundations of decision theory." *Australasian Journal of Philosophy* 89:641–63.

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* 21:1087–92.

Michaelian, Kourken. 2011. "The epistemology of forgetting." *Erkenntnis* 74:399–424.

Millar, Alan. 2011. "Why knowledge matters." *Proceedings of the Aristotelian Society* 85:63–81.

Millson, Jared. 2021. "Seeking confirmation: A puzzle for norms of inquiry." *Analysis* 80:683–93.

Miracchi, Lisa. 2019. "When evidence isn't enough: Suspension, evidentialism, and knowledge-first virtue epistemology." *Episteme* 16:413–37.

Misak, Cheryl. 1987. "Peirce, Levi, and the aims of inquiry." *Philosophy of Science* 54:256–65.

Morton, Jennifer. 2017. "Reasoning under scarcity." *Australasian Journal of Philosophy* 95:543–59.

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

Muñoz, Daniel. 2021. "The rejection of consequentializing." *Journal of Philosophy* 118:79–96.

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.

Munton, Jessie. 2019. "Beyond accuracy: Epistemic flaws with statistical generalizations." *Philosophical Issues* 29:228–40.

Murdock, Bennet. 1962. "The serial position effect of free recall." *Journal of Experimental Psychology* 64:482–8.

Mussweiler, Thomas and Strack, Fritz. 1997. "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility." *Journal of Personality and Social Psychology* 73:437–46.

Mussweiler, Thomas and Strack, Fritz. 1999. "Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model." *Journal of Experimental Social Psychology* 35:136–64.

Mussweiler, Thomas and Strack, Fritz. 2000. "Comparing is believing: A selective accessibility model of judgmental anchoring." In Wolfgang Stroebe and Miles Hewstone (eds), *European Review of Social Psychology*, volume 10, 135–67. Wiley.

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* 53:431–44.

Newell, Allen, Shaw, Cliff, and Simon, Herbert. 1956. "Chess playing programs and the problem of complexity." *IBM Journal of Research and Development* 4:320–35.

Newell, Allen and Simon, Herbert. 1965. "An example of human chess play in light of chess playing programs." In Norbert Wiener and J.P. Schadé (eds), *Progress in Biocybernetics*, volume 2, 19–75. Elsevier.

Newton, Lilith. manuscript. "Doubt as the inquiry starter." Manuscript.

Nilsson, Hakan, Rieskamp, Jörg, and Jenny, Mirjam A. 2013. "Exploring the overestimation of conjunctive probabilities." *Frontiers in Psychology* 4:101.

Nisbett, Richard, Borgida, Eugene, Crandall, Rick, and Reed, Harvey. 1976. "Popular induction: Information is not necessarily informative." In John Carroll and John Payne (eds), *Cognition and social behavior*. Lawrence Erlbaum.

Norcross, Alastair. 1997. "Consequentialism and commitment." *Pacific Philosophical Quarterly* 78:380–403.

Northcraft, Gregory and Neale, Margaret. 1987. "Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions." *Organizational Behavior and Human Decision Processes* 39:84–97.

Oaksford, Mike and Chater, Nick. 1994. "A rational analysis of the selection task as optimal data selection." *Psychological Review* 101:608–31.

Oaksford, Mike and Chater, Nick. 2003. "Conditional probability and the cognitive science of conditional reasoning." *Mind and Language* 18:359–79.

Oaksford, Mike and Chater, Nick. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Oaksford, Mike and Chater, Nick. 2020. "New paradigms in the psychology of reasoning." *Annual Review of Psychology* 71:305–30.

Oaksford, Mike, Chater, Nick, and Grainger, Becki. 1999. "Probabilistic effects in data selection." *Thinking and Reasoning* 5:193–243.

Oaksford, Mike, Chater, Nick, Grainger, Becki, and Larkin, Joanne. 1997. "Optimal data selection in the reduced array selection task (RAST)." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:441–58.

Oberauer, Klaus, Wilhelm, Oliver IV, and Rosas Diaz, Ricardo. 1999. "Bayesian rationality for the Wason selection task? A test of optimal data selection theory." *Thinking and Reasoning* 5:115–44.

Olsson, Erik J. and Westlund, David. 2006. "On the role of the research agenda in epistemic change." *Erkenntnis* 65:165–83.

Oskamp, Stuart. 1965. "Overconfidence in case-study judgments." *Journal of Consulting Psychology* 29:261–5.

Pachur, Thorsten, Gigerenzer, Gerd, Schooler, Lael, and Goldstein, Daniel. 2012. "When is the recognition heuristic an adaptive tool?" In Peter Todd and Gerd Gigerenzer (eds), *Ecological rationality: Intelligence in the world*, 113–43. Oxford University Press.

Pachur, Thorsten, Todd, Peter, Gigerenzer, Gerd, Schooler, Lael, and Goldstein, Daniel. 2011. "The recognition heuristic: A review of theory and tests." *Frontiers in Psychology* 2:1–14.

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

Palmira, Michele. 2020. "Inquiry and the doxastic attitudes." *Synthese* 197:4947–73.

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* 135:531–54.

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.

Paul, L.A. 2014. *Transformative experience*. Oxford University Press.

Peirce, C.S. 1877. "The fixation of belief." *Popular Science Monthly* 12:1–15.

Perham, Nick and Oaksford, Mike. 2005. "Deontic reasoning with emotional content: Evolutionary psychology or decision theory?" *Cognitive Science* 29:681–718.

Pettigrew, Richard. 2012. "Accuracy, chance, and the principal principle." *Philosophical Review* 121:241–75.

Pettigrew, Richard. 2013. "A new epistemic utility argument for the principal principle." *Episteme* 10:19–35.

Pettigrew, Richard. 2016. "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research* 92:35–59.

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

Pettit, Philip. 1989. "Consequentialism and respect for persons." *Ethics* 100:116–26.

Pettit, Philip and Smith, Michael. 2000. "Global consequentialism." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds), *Morality, rules and consequences*, 121–33. Edinburgh University Press.

Piaget, Jean. 1957. *Logic and psychology*. Basic books.

Pierson, Emma, Simoiu, Camelia, Overgoor, Jan, Corbett-Davies, Sam, Jenson, Daniel, Shoemaker, Amy, Ramachandran, Vignesh, Barghouty, Phoebe, Phillips, Cheryl, Shroff, Ravi, and Goel, Sharad. 2020. "A large-scale analysis of racial disparities in police stops across the United States." *Nature Human Behavior* 4:736–45.

Plantinga, Alvin. 1983. "Reason and belief in God." In Alvin Plantinga and Nicholas Wolsteroff (eds), *Faith and rationality: Reason and belief in God*, 16–93. Notre Dame University Press.

Podgorski, Abelard. 2017. "Rational delay." *Philosophers' Imprint* 17:1–19.

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

Preston-Roedder, Ryan. 2013. "Faith in humanity." *Philosophy and Phenomenological Research* 87:664–87.

Proust, Joëlle. 2013. *The philosophy of metacognition*. Oxford University Press.

Railton, Peter. 1984. "Alienation, consequentialism, and the demands of morality." *Philosophy and Public Affairs* 13:134–71.

Ramsey, Frank. 1926/1931. "Truth and probability." In R.B. Braithwaite (ed.), *The foundations of mathematics and other logical essays*. Harcourt.

Reyna, Valerie and Brainerd, Charles. 2008. "Numeracy, ratio bias, and denominator neglect in judgments of risk and probability." *Learning and individual differences* 18:89–107.

Richter, Reed. 1990. "Ideal rationality and hand waving." *Australasian Journal of Philosophy* 68:147–56.

Rieskamp, Jörg and Reimer, Torsten. 2007. "Ecological rationality." In Roy F. Baumeister and Kathleen Vohs (eds), *Encyclopedia of social psychology*, 273–4. Sage.

Rinard, Susanna. 2015. "Against the new evidentialists." *Philosophical Issues* 25:208–23.

Rinard, Susanna. 2017. "No exception for belief." *Philosophy and Phenomenological Research* 94:121–43.

Rinard, Susanna. 2019a. "Believing for practical reasons." *Noûs* 53:763–84.

Rinard, Susanna. 2019b. "Equal treatment for belief." *Philosophical Studies* 176:1923–50.

Rips, Lance. 1994. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press.

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* 32:131–70.

Ross, Lee, Lepper, Mark, and Hubbard, Michael. 1975. "Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm." *Journal of Personality and Social Psychology* 32:880–92.

Rubinstein, Ariel. 1989. "The electronic mail game: Strategic behavior under 'almost common knowledge.'" *American Economic Review* 79:385–91.

Rubinstein, Ariel. 1998. *Modeling bounded rationality*. MIT Press.

Rubinstein, Ariel. 2012. *Economic fables*. OpenBook Publishers.

Russell, Stewart and Wefald, Eric. 1991. "Principles of metareasoning." *Artificial Intelligence* 49:361–95.

Samuelson, Paul A. 1938. "A note on the pure theory of consumer's behaviour." *Economica* 5:61–71.

Savage, Leonard. 1954. *The foundations of statistics*. Wiley.

Savage, Leonard. 1972. *The foundations of statistics*. Courier Corporation.

Scheffler, Samuel. 1994. *The rejection of consequentialism*. Oxford University Press.

Schervish, Mark, Seidenfeld, Teddy, and Kadane, Joseph. 2000. "How sets of coherent probabilities may serve as models for degrees of incoherence." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 8:347–55.

Schervish, Mark, Seidenfeld, Teddy, and Kadane, Joseph. 2002. "Measuring Incoherence." *Sankhya: The Indian Journal of Statistics* 64:561–87.

Schervish, Mark, Seidenfeld, Teddy, and Kadane, Joseph. 2003. "Measures of incoherence: How not to gamble if you must." In José Bernardo (ed.), *Bayesian statistics*, 385–401. Oxford University Press.

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* 140:921–48.

Schooler, Lael J. and Hertwig, Ralph. 2005. "How forgetting aids heuristic inference." *Psychological Review* 112:610–28.

Schroeder, Mark. 2007. "Teleology, agent-relative value, and 'good.'" *Ethics* 117:265–95.

Schroeder, Mark. 2018. "When beliefs wrong." *Philosophical Topics* 46:115–28.

Schroyens, Walter, Schaeken, Walter, and d'Ydewalle, Géry. 2001. "The processing of negations in conditional reasoning: A meta-analytic case study in mental model and/or mental logic theory." *Thinking and Reasoning* 7:121–72.

Sedlmeier, Peter and Gigerenzer, Gerd. 2001. "Teaching Bayesian reasoning in less than two hours." *Journal of Experimental Psychology: General* 130:380–400.

Seidel, Christian. 2019. "New wave consequentialism: An introduction." In Christian Seidel (ed.), *Consequentialism: New directions, new problems*, 1–30. Oxford University Press.

Selten, Reinhard. 1998. "Aspiration adaptation theory." *Journal of Mathematical Psychology* 42:191–214.

Setiya, Kieran. 2018. "Must consequentialists kill?" *Journal of Philosophy* 115:92–105.

Shah, Anuj, Mullainathan, Sendhil, and Shafir, Eldar. 2012. "Some consequences of having too little." *Science* 338:682–5.

Shah, Nishi. 2003. "How truth governs belief." *Philosophical Review* 112:447–82.

Shah, Nishi. 2006. "A new argument for evidentialism." *Philosophical Quarterly* 56:481–98.

Shenhav, Amitai, Botvinick, Matthew, and Cohen, Jonathan. 2013. "The expected value of control: An integrative theory of anterior cingulate cortex function." *Neuron* 79:217–40.

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.

Shimony, Soloman Eyal. 1994. "Finding MAPs for belief networks is NP-hard." *Artificial Intelligence* 68:399–410.

Siegel, Susanna. 2017. *The rationality of perception*. Oxford University Press.

Silk, Alex. 2014. "Evidence sensitivity in weak necessity deontic modals." *Journal of Philosophical Logic* 43:691–723.

Silver, David, Hubert, Thomas, Schrittwieser, Julian, Antonoglou, Ioannis, Lai, Matthew, Guez, Arthur, Lanctot, Marc, Sifre, Laurent, Kumaran, Dharshan, Graepel, Thore, Lillicrap, Timothy, Simonyan, Karen, and Hassabis, Demis. 2018. "A general reinforcement-learning algorithm that masters chess, shogi, and Go through self-play." *Science* 362:1140–4.

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* 99:917–32.

Simon, Herbert. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics* 69:99–118.

Simon, Herbert. 1959. *Models of man*. Wiley.

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

Simon, Herbert. 1990. "Invariants of human behavior." *Annual Review of Psychology* 41:1–20.

Simon, Herbert. 1998. "Letter to Ariel Rubinstein." In Ariel Rubinstein (ed.), *Modeling bounded rationality*, 187–90. MIT Press.

Simon, Herbert and Gilmartin, Kevin J. 1979. "A simulation of memory for chess positions." *Cognitive Psychology* 5:29–46.

Sims, Christopher. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50:665–90.

Singmann, Henrik and Klauer, Karl Christoph. 2011. "Deductive and inductive conditional inferences: Two modes of reasoning." *Thinking and Reasoning* 17:247–81.

Singmann, Henrik, Klauer, Karl Christoph, and Sieghard, Beller. 2016. "Probabilistic conditional reasoning: Disentangling form and content with the dual-source model." *Cognitive Psychology* 88:61–87.

Siscoe, Robert Weston. 2021. "Belief, rational and justified." *Mind* 130:59–83.

Slote, Michael. 1984. "Satisficing consequentialism." *Proceedings of the Aristotelian Society, Supplement* 58:139–63.

Slovic, Paul, Finucane, Melissa, Peters, Ellen, and MacGregor, Donald. 2007. "The affect heuristic." *European Journal of Operational Research* 177:1333–52.

Slovic, Paul and Tversky, Amos. 1974. "Who accepts Savage's axiom?" *Behavioral Science* 19:368–73.

Smith, Angela. 2005. "Responsibility for attitudes: Activity and passivity in mental life." *Ethics* 115:236–71.

Smith, Holly. 2014. "The subjective moral duty to inform oneself before acting." *Ethics* 125:11–38.

Smithies, Declan. 2015. "Ideal rationality and logical omniscience." *Synthese* 192:2769–93.

Sober, Elliott. 1978. "Psychologism." *Journal of Social Behavior* 8:165–91.

Staffel, Julia. 2015. "Measuring the overall incoherence of credence functions." *Synthese* 192:1467–93.

Staffel, Julia. 2017. "Graded incoherence for accuracy-firsters." *Philosophy of Science* 84:189–213.

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

Stalnaker, Robert. 1981. "Letter to David Lewis, May 21, 1972." In William Harper, Robert Stalnaker, and Glenn Pearce (eds), *Ifs: Conditionals, belief, decision, chance, and time*, 151–2. Reidel.

Stalnaker, Robert. 1984. *Inquiry*. MIT Press.

Stanovich, Keith. 1999. *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.

Stanovich, Keith and West, Richard. 2000. "Individual differences in reasoning: Implications for the rationality debate?" *Behavioral and Brain Sciences* 23:645–65.

Steffensmeier, Darrell and Demuth, Stephen. 2000. "Ethnicity and sentencing outcomes in US federal courts: Who is punished more harshly?" *American Sociological Review* 65:705–29.

Steglich-Petersen, Asbjørn. 2011. "How to be a teleologist about epistemic reasons." In Andreas Reisner and Asbjørn Steglich-Petersen (eds), *Reasons for belief*, 13–33. Cambridge University Press.

Steglich-Petersen, Asbjørn. forthcoming. "An instrumentalist unification of zetetic and epistemic reasons." *Inquiry* forthcoming.

Steglich-Petersen, Asbjørn and Skipper, Matthias. 2019. "An instrumentalist account of how to weigh epistemic and practical reasons for belief." *Mind* 129:1071–94.

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

Steinberger, Florian. 2019. "Consequence and normative guidance." *Philosophy and Phenomenological Research* 98:306–28.

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* 69:213–25.

Stocker, Michael. 1976. "The schizophrenia of modern ethical theories." *Journal of Philosophy* 73:453–66.

Striker, Gisela. 2001. "Skepticism as a kind of philosophy." *Archiv für Geschichte der Philosophie* 83:113–29.

Stroud, Sarah. 2006. "Epistemic partiality in friendship." *Ethics* 116:498–524.

Sturm, Thomas. 2019. "Formal versus bounded norms in the psychology of rationality: Toward a multilevel analysis of their relationship." *Philosophy of the Social Sciences* 49:190–209.

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

Sylvan, Kurt. manuscript. *Respect for truth: A Kantian account of theoretical normativity*.

Tenenbaum, Sergio. 2014. "The perils of honest consequentializing." *Philosophy and Phenomenological Research* 88:233–40.

Thagard, Paul and Nisbett, Richard. 1983. "Rationality and charity." *Philosophy of Science* 50:250–67.

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

Thorstad, David. 2021. "Inquiry and the epistemic." *Philosophical Studies* 178:2913–28.

Thorstad, David. 2022a. "There are no epistemic norms of inquiry." *Synthese* 200:410.

Thorstad, David. 2022b. "Two paradoxes of bounded rationality." *Philosophers' Imprint* 22:1–16.

Thorstad, David. forthcoming. "The accuracy-coherence tradeoff in cognition." *British Journal for the Philosophy of Science* forthcoming.

Thorstad, David. forthcoming b. "General-purpose institutional decisionmaking heuristics: The case of decisionmaking under deep uncertainty." *British Journal for the Philosophy of Science* .

Thorstad, David. forthcoming c. "Why bounded rationality (in epistemology)?" *Philosophy and Phenomenological Research* forthcoming.

Thorstad, David. manuscript a. "The dual deontic role of unpossessed evidence."

Thorstad, David. manuscript b. "Norms of inquiry."

Thorstad, David. manuscript c. "Inquiry under bounds."

Todd, Peter, Fiddick, Laurence, and Krauss, Stefan. 2000. "Ecological rationality and its contents." *Thinking and Reasoning* 6:375–84.

Todd, Peter and Gigerenzer, Gerd. 2012a. *Ecological rationality: Intelligence in the world.* Oxford University Press.

Todd, Peter and Gigerenzer, Gerd. 2012b. "What is ecological rationality?" In Peter Todd and Gerd Gigerenzer (eds), *Ecological rationality: Intelligence in the world*, 3–30. Oxford University Press.

Turner, Brandon and Schley, Dan. 2016. "The anchor integration model: A descriptive model of anchoring effects." *Cognitive Psychology* 90:1–47.

Tversky, Amos. 1969. "Intransitivity of preferences." *Psychological Review* 76:31–48.

Tversky, Amos and Kahneman, Daniel. 1971. "Belief in the law of small numbers." *Psychological Bulletin* 76:105–10.

Tversky, Amos and Kahneman, Daniel. 1973. "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology* 5:207–32.

Tversky, Amos and Kahneman, Daniel. 1974. "Judgment under uncertainty: Heuristics and biases." *Science* 185:1124–31.

Tversky, Amos and Kahneman, Daniel. 1981. "The framing of decisions and the psychology of choice." *Science* 211:453–8.

Tversky, Amos and Kahneman, Daniel. 1982. "Judgments of and by representativeness." In Daniel Kahneman, Paul Slovic, and Amos Tversky (eds), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Tversky, Amos and Kahneman, Daniel. 1983. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review* 90:293–315.

Tversky, Amos and Kahneman, Daniel. 1986. "Rational Choice and the Framing of Decisions." *Journal of Business* 59:S251–78.

Tversky, Amos and Koehler, Derek. 1994. "Support theory: A nonextensional representation of subjective probability." *Psychological Review* 101:547–67.

Vindrola, Filippo and Crupi, Vincenzo. forthcoming. "Bayesians too should follow Wason: A comprehensive accuracy-based analysis of the selection task." *British Journal for the Philosophy of Science* forthcoming.

Vohs, Kathleen. 2013. "The poor's poor mental power." *Science* 341:969–70.

Von Neumann, John and Morgenstern, Oskar. 1947. *Theory of games and economic behavior.* Princeton University Press, second edition.

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.

Wason, Peter. 1966. "Reasoning." In Brian Foss (ed.), *New horizons in psychology*, 135–51. Penguin.

Wason, Peter C. 1968. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology* 20:273–81.

Way, Jonathan. 2012. "Transmission and the wrong kind of reason." *Ethics* 122:489–515.

Wedgwood, Ralph. 2018. "Epistemic teleology: Synchronic and diachronic." In Kristoffer Ahlström-Vij and Jeffrey Dunn (eds), *Epistemic consequentialism*, 85–112. Oxford University Press.

Weinberg, Jonathan, Nichols, Shaun, and Stich, Stephen. 2001. "Normativity and epistemic intuitions." *Philosophical Topics* 29:429–60.

Weisberg, Jonathan. 2015. "You've come a long way, Bayesians." *Journal of Philosophical Logic* 44:817–34.

Whitcomb, Dennis. 2010. "Curiosity was framed." *Philosophy and Phenomenological Research* 81:664–87.

Whitcomb, Dennis, Battaly, Heather, Baehr, Jason, and Howard-Snyder, Daniel. 2017. "Intellectual humility: Owning our limitations." *Philosophy and Phenomenological Research* 44:509–39.

Whitt, Laurie Anne. 1992. "Indicators of theory promise." *Philosophy of Science* 59:612–34.

Wiens, David. 2020. "The general theory of the second best is more general than you think." *Philosophers' Imprint* 5:1–26.

Williams, Bernard. 1973. *Problems of the self*. Cambridge University Press.

Williamson, Timothy. 2000. *Knowledge and its limits*. Oxford University Press.

Wilson, Timothy, Houston, Christopher, Etling, Kathryn, and Brekke, Nancy. 1996. "A new look at anchoring effects: Basic anchoring and its antecedents." *Journal of Experimental Psychology: General* 125:387–402.

Woodard, Elise. forthcoming. "Why double-check?" *Episteme* forthcoming.

Woodard, Elise. forthcoming a. "The ignorance norm and paradoxical assertions." *Philosophical Topics* .

Woodard, Elise. forthcoming b. "Why double-check?" *Episteme* forthcoming.

Woodard, Elise and Flores, Carolina. forthcoming. "Epistemic vigilance: In defense of epistemic norms on evidence-gathering." *Philosophical Studies* forthcoming.

Worsnip, Alex. 2021. *Fitting things together: Coherence and the demands of structural rationality*. Oxford University Press.

Yablo, Stephen. 2014. *Aboutness*. Princeton University Press.

Yalcin, Seth. 2007. "Epistemic modals." *Mind* 116:983–1026.

Yalcin, Seth. 2016. "Modalities of normativity." In Nate Charlow and Matthew Chrisman (eds), *Deontic modality*, 230–55. Oxford University Press.

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

Zajonc, Robert B. 1968. "Attitudinal effects of mere exposure." *Journal of Personality and Social Psychology* 9:1–27.

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*, 270–89. Springer.

Zynda, Lyle. 1996. "Coherence as an ideal of rationality." *Synthese* 109:175–216.

Zynda, Lyle. 2000. "Representation theorems and realism about degrees of belief." *Philosophy of Science* 67:45–69.

Index