Idealization and bounded rationality

Abstract

Debates about bounded rationality often appear to turn on the degree of permissible idealization in modeling. I argue that many such debates have been mislocated. The primary object of contention is not the degree of permissible idealization nor even the criteria for permissible idealization, but rather whether shared criteria for permissible idealization justify proposed idealizations involving cognitive bounds. I support this diagnosis through a pair of schematic arguments and a pair of case studies from prominent defenses of idealization in modeling. I conclude with lessons for the distinction between ideal and non-ideal epistemology, the justificatory burdens of idealization and de-idealization, and the prospects for Bayesian approaches to bounded rationality.

1 Introduction

Humans are bounded agents. We are limited internally by our cognitive architecture, and externally by our environment. The study of bounded rationality asks what rationality requires of bounded agents (Icard 2018; Gigerenzer and Selten 2001; Thorstad 2024b).

Debates about bounded rationality often appear to turn on:

> **(Degree of Permissible Idealization)** For any subject S and purpose P, how idealized should models of S for purpose P be?

Critics of bounded rationality remind us of the importance of idealization and link their project to a defense of idealization:

> Idealization is ubiquitous and inescapable in epistemology. As such, there isn't a feasible, general project of "de-idealizing" the field. ... We're all ideal theorists, and inevitably so. (Greco 2023, p. 135)

> Without an idealized epistemology ... all we can expect to produce is a more-or-less-detailed report on our actual linguistic practices — epistemology as lexicology. (Pasnau 2013, p. 989)

> Ideal epistemologists are concerned with questions about what perfectly rational, cognitively idealized, computationally unlimited believers would believe. (Carr 2022, p. 1132)

Defenders of bounded rationality remind us of the pitfalls of idealization and link their project to a reduction in idealization:

> 'Ideal epistemology'—the kind of epistemology that works with … idealized models—goes wrong in that we sometimes need to work with less idealized models of human beings and of the social interactions between them. (McKenna 2023, p. 3)

> Ideal Epistemology is a research programme that idealizes away from cognitive limitations and studies norms of rational thinking for cognitively ideal, or close to ideal, agents … Given that we can't think like cognitively ideal agents, how should we think? Answering this question is the job of a theory of non-ideal (aka bounded) epistemic rationality. (Hughes 2024, pp. 73-4)

My core thesis is that many debates which seem to turn on Degree of Permissible Idealization instead turn on whether proposed idealizations meet shared standards for permissible idealization.

In key cases of interest, I argue that bounded rationality theorists can establish the Misapplication Thesis that idealizations do not meet, or have not been shown to meet shared standards for permissible idealization. Establishing the Misapplication Thesis allows bounded rationality theorists to reject proposed idealizations while replacing them with models that are in many cases no less idealized than the rejected models. More generally, establishing the Misapplication Thesis across specific debates suggests that these debates about bounded rationality have been mislocated. What appears to be an entrenched general debate about the role of idealization is often a local and tractable debate about the justifiability of proposed idealizations according to shared standards.

Here is the plan. Section 2 clarifies and motivates the Misapplication Thesis. Sections 3-4 defend the Misapplication Thesis in two case studies. The first looks at Milton Friedman's (1953) classic defense of idealization in "The methodology of positive economics". Bounded rationality began as a reaction against neoclassical economics (Klaes and Sent 2005; Simon 1947), and we will see here that the Misapplication Thesis gets a definite take against the best-known neoclassical defense of idealization. The second case study looks at a recent epistemological defense of idealization by Daniel Greco (2023). Again, we will see that Greco's modest modeling approach has not been shown to ground a key idealization, demonstrating the continued relevance of the Misapplication Thesis to contemporary epistemological debates. Section 5 concludes with implications for the distinction between ideal and non-ideal epistemology, the justificatory burdens of idealization and de-idealization, and the prospects for Bayesian approaches to bounded rationality.

2 Preliminaries

2.1 Clarifying the target

Two distinctions will help to clarify the type of idealization at issue in the Misapplication Thesis. First, idealizations can be *normative* or *descriptive* (Carr 2022; Mills 2005; Schmidtz 2011). Normative idealizations seek to directly characterize a normative object, such as a best or most perfect state or a regulative ideal. In this sense, it may be claimed that a normatively ideal epistemic agent would be unbounded (Carr 2022) or that the best credence functions are fully coherent (Staffel 2020). By contrast, descriptive idealizations do not seek to directly characterize a normative object, but instead distort or omit aspects of a situation for predictive or explanatory purposes such as increasing model tractability (McMullin 1985) or drawing attention to the most important explanatory features of a situation (Cartwright 1989).[1] This paper defends the Misapplication Thesis as a claim

[1]See Elliott-Graves and Weisberg (2014) for a review of descriptive idealization.

about descriptive rather than normative idealizations.

Second, we need to distinguish *cognitive bounds* such as limited cognitive abilities and computational costs from other objects of idealization. For example, debates about idealization in political philosophy often turn on the appropriateness of Rawls' assumptions that society is sufficiently developed to meet the demands of justice and that agents fully comply with these demands (Mills 2005; Schmidtz 2011; Simmons 2010). We will see in Section 2.3 that Rawlsians do not deny the importance of modeling cognitive bounds. This paper defends the Misapplication Thesis as a claim about descriptive idealizations involving human cognitive bounds, but not about further descriptive idealizations.

2.2 Formulating the Misapplication Thesis

We saw in Section 1 that debates about bounded rationality often seem to turn on:

> **(Degree of Permissible Idealization)** For any subject S and purpose P, how idealized should models of S for purpose P be?

These discussions often evolve into downstream debates about the criteria for permissible idealization. Thus, an alternative locus of disagreement might be:

> **(Criteria for Permissible Idealization)** For any subject S, purpose P, and idealization I, what determines the appropriateness of I in models of S for purpose P?

My central claim is that many debates about bounded rationality can be resolved without settling disputes over Degree of Permissible Idealization or Criteria for Permissible Idealization.

This can be done by establishing:

> **(Misapplication Thesis)** For some subject S, purpose P and descriptive idealization I involving cognitive bounds, I does not meet, or has not been shown to

meet the criteria for permissible idealization accepted by the theorist proposing *I*.

When the Misapplication Thesis holds, proposed idealizations can be rejected without resolving disagreements, if any, over Criteria for Permissible Idealization.

Further, we will see that in many cases proposed counter-models are no less idealized than the rejected models. In this case, Degree of Permissible Idealization also does not capture the main locus of disagreement between bounded rationality theorists and their opponents. I return to this point in Section 5.1.

2.3 Motivating the Misapplication Thesis

Before beginning, I want to offer two preliminary arguments for thinking that the Misapplication Thesis may often supplant Degree of Permissible Idealization and Criteria for Permissible Idealization as the key object of contention between bounded rationality theorists and their opponents.

The first is the Argument from Bounded Idealization. The Argument from Bounded Idealization begins by noting that many bounded rationality models are deeply idealized, often more idealized than many unbounded rationality models are. Most directly, the tradition of Bayesian rational analysis (Anderson 1990; Chater and Oaksford 1999b; Icard 2018) employs sophisticated Bayesian probabilistic and decision-theoretic models to explain the behavior of agents under constraints such as limited computational power (Icard ms), reasoning ability (Oaksford and Chater 2007), or memory (Anderson 1990). This tradition idealizes to such a degree that one of the main complaints raised by critics continues to be that humans could not possibly cognize in the way set out by Bayesian models (Bowers and Davis 2012; Holton 2014; Jones and Love 2011). Outside of the Bayesian tradition, we find highly idealized and mathematically sophisticated treatments of topics such as tradeoffs (Johnson and Payne 1985; Thorstad 2024a, 2025; Wheeler 2020) and the performance of specific heuristics across environments (Dawes and Corrigan 1974;

Dawes 1979; Hogarth and Karelaia 2006; Martignon 2001).

The Argument from Bounded Idealization asks what best explains the high degree of idealization found in many bounded rationality models. The Argument from Bounded Idealization suggests that a good explanation of the high degree of idealization in many bounded rationality models is that bounded rationality theorists are not driven by a principled opposition to idealization, even when that idealization involves complex mathematics or ignoring of some cognitive bounds in order to highlight others. Rather, bounded rationality theorists are driven by the explanatory importance of the cognitive bounds that they model. When bounded rationality theorists insist that models of some target phenomenon should incorporate a particular cognitive bound, this insistence is often driven by the explanatory importance of this particular bound in this particular situation, rather than by a general opposition to idealization.

A second preliminary motivation for the Misapplication Thesis is the Argument from Ideal Variation. This argument notes that self-styled ideal theorists in different domains often make different idealizations. In particular, while epistemologists and moral philosophers often idealize away from cognitive bounds, the debates about ideal theory in political philosophy introduced in Section 2.1 do not typically turn on cognitive bounds. Rawls himself was abundantly clear that he did not take ideal theory to abstract away from cognitive bounds, and in fact took human cognitive bounds to provide a strike against theories such as utilitarianism (Simmons 2010).

The Argument from Ideal Variation asks what best explains the fact that some ideal theorists incorporate cognitive bounds while others do not. The Argument from Ideal Variation suggests that a good preliminary explanation of this phenomenon is that ideal theorists are not moved by general commitments about idealization which rule out modeling cognitive bounds. Rather, they think that in some specific domains and for some purposes, idealizing away from cognitive bounds is an appropriate idealization.

The Argument from Bounded Idealization and the Argument from Ideal Variation offer preliminary practice-driven reasons to think that the disagreement between bounded

rationality theorists and their opponents may often turn not on Degree of Permissible Idealization nor even on Criteria for Permissible Idealization, but rather the Misapplication Thesis that shared criteria for permissible idealization do not justify a given descriptive idealization away from cognitive bounds. But these are only preliminary motivations, and they tell us only why the Misapplication Thesis is likely to be at issue, not why it is often true. This paper defends the Misapplication Thesis in case studies of racial discrimination in hiring (Section 3) and the simplifying role of full belief in human cognition (Section 4).

3 Friedman on perfect competition and hiring discrimination

Bounded rationality began as a reaction against neoclassical economics (Klaes and Sent 2005; Simon 1947). A good place to begin is by showing how the Misapplication Thesis gets a take against one of the defining methodological papers of the neoclassical approach.

3.1 Friedman's defense of idealization

Milton Friedman's (1953) "The methodology of positive economics" has been called the most influential discussion of economic methodology in the twentieth century (Mäki 2009). It is also one of the best-known defenses of idealization in modeling.

Friedman distinguishes *positive economics*, which deals with descriptive questions about what is, from *normative economics*, which deals with normative questions about what ought to be. Friedman argues that positive economics should often make extensive use of idealization.

Friedman presents many classic arguments in favor of idealization. The most famous has come to be known as the F-Twist (Musgrave 1981; Wong 1973):

> **(F-Twist)** Models are successful when agents behave as if they obeyed model assumptions, whether or not those assumptions hold.

More generally, Friedman argues that:

> **(Modeling Costs)** De-idealized models are often costly to construct and apply.
>
> **(Theoretical Virtues)** Idealized models often exhibit theoretical virtues such as simplicity and fruitfulness.
>
> **(Competitive Evaluation)** Models should not be rejected until opponents can supply a better model.

Much ink has been spilled about Friedman's general arguments for idealization (Mäki 2009; McCloskey 1985; Frazer and Boland 1983). But what is not always emphasized is what Friedman wants to do with these arguments.

Friedman's main stalking horse is the theory of imperfect competition as developed by Edward Chamberlin (1923) and Joan Robinson (1933). Immediately after his initial defense of idealization, Friedman launches an extended attack on imperfect competition, whose faults he takes to be almost immediately revealed by his defense of idealization. Friedman argues that "the theory of monopolistic and imperfect competition is one example of the neglect in economic theory of these propositions" about idealization (Friedman 1953, p. 153), and that "it is only a short step" from his examples "to the economic hypothesis that under a wide range of circumstances individual firms" behave as the theory of perfect competition predicts (Friedman 1953, p. 157).

Friedman applies this analysis to a case study of racial discrimination in hiring. Friedman notes that many sociologists may take themselves to have excellent data to suggest that employers in the 1950s racially discriminated in hiring on the basis of managerial prejudice.[2] However, Friedman suggests, many markets are well-modeled as perfectly competitive and perfect competition should eliminate racial discrimination driven by managerial prejudice. As a result, Friedman suggests, without engaging in any extensive discussion of the sociologist's data nor in modeling of any kind, many economists will

[2]See Pager and Shepherd (2008) for a review.

be considerably more skeptical about models which incorporate racial discrimination in hiring.

3.2 When perfect competition eliminates hiring discrimination

Friedman does not elaborate on his reasons for taking perfect competition to eliminate hiring discrimination. While the best response would be to ask Friedman to elaborate on his reasons, many economists at the time were drawn to an argument soon to be popularized by Kenneth Arrow (1972; 1973). Let us begin with this leading argument and see how far it takes us in eliminating hiring discrimination.

In the 1950s, the reigning economic model of hiring discrimination was Gary Becker's (1957) account of taste-based discrimination.[3] On this view, discrimination is driven by direct distaste for association with minorities.

One of Becker's models assumes that employers themselves dislike hiring minorities. In this model, workers are divided into two groups: a favored group A and a disfavored group B. Under perfect competition, these groups receive equilibrium wages w_A, w_B and are hired by identically-sized firms in numbers L_A, L_B per firm. Workers have identical marginal productivity f. Firms produce outputs according to a concave production function $F(L_A + L_B)$ and sell them at fixed price p.

In this model, firm profits $\pi(L_A, L_B)$ are the sale price of outputs less wages paid, so that:

$$\pi(L_A, L_B) = pF(L_A + L_B) - w_A L_A - w_B L_B. \tag{1}$$

However, firms dislike hiring B workers, experiencing disutility $d > 0$ for each B worker they hire. As a result, firms maximize a value function V with

[3]Note that Friedman (1953) predates the publication of Becker's (1957) work by four years. As a result, while Friedman's argument is very much in the spirit of Becker's first model of taste-based discrimination, I do not mean to assert that Friedman intended exactly this argument. This is simply the most natural and least anachronistic way to gloss the kind of argument that Friedman might have supplied if he had given a detailed argument for his views on hiring discrimination.

$$V(L_A, L_B) = \pi(L_A, L_B) - dL_B. \tag{2}$$

In a perfectly competitive equilibrium, $dV/dL_A = dV/dL_B = 0$, forcing wages to $w_A = pf$ and $w_B = pf/(1+d)$. In this equilibrium, A workers are paid the full market value of their marginal product, but B workers receive only a fractional share of the market value of their product, with employers keeping the rest in compensation for their dislike of hiring B workers.

A common response to this model popularized by Kenneth Arrow (1972; 1973) is that the discriminatory equilibrium is unstable. A new firm entering the market which does not dislike hiring B workers would be able to replace A workers with B workers and pocket the difference $pfd/(1+d)$ in wages. Over time, this new business model would drive prices down, bankrupting discriminatory firms until only non-discriminatory firms remained. In this way, Becker's first model of taste discrimination makes long-run hiring discrimination impossible under perfect competition, as Friedman claims.

3.3 Modeling hiring discrimination

At this point, three responses are in order. First, there are known ways to restore hiring discrimination under perfect competition in taste-based discrimination models (Section 3.3.1). Second, there are promising alternative explanations of hiring discrimination under perfect competition beyond taste-based discrimination (Section 3.3.2). Finally, there is good evidence to suggest that some hiring discrimination emerges from failures of perfect competition which can be tractably modeled and studied (Section 3.3.3).

In support of the Misapplication Thesis, all models used to make these responses perform well on the criteria used by Friedman to motivate idealization. In this way, a surface dispute about the need for idealization in modeling is here best understood as an under-evidenced attempt to defend inadequate models which do not meaningfully engage with data supporting hiring discrimination.

3.3.1 Restoring equilibrium taste-based discrimination

Becker (1957) gave two further models of taste-based discrimination, both of which allow discrimination to persist in a perfectly competitive equilibrium. The simplest model involves consumer discrimination.

In this model, firms are unprejudiced so that $d = 0$ in (2). However, consumers are prejudiced, preferring to purchase goods from segregated firms that do not employ B workers. They thus experience disutility $d' > 0$ for each good purchased from an unsegregated firm. Consuming x_A units of goods produced from segregated firms and x_B units of goods produced by unsegregated firms thus provides utility:

$$U(x_A, x_B) = x_A + (1 - d')x_B. \tag{3}$$

This forces the equilibrium prices p_A, p_B of goods produced by segregated and unsegregated firms to obey the relationship $p_A = p_B + d'$. Under perfect competition, workers are paid the full market value of their output so that $w_A = fp_A$ and $w_B = fp_B$. Because $p_A = p_B + d'$, this means that A workers earn fd' more than B workers for their labor.

None of Friedman's arguments for idealization tell against such models. In agreement with the F-Twist, modern economists defend consumer discrimination models on the grounds that they improve predictive accuracy (Borjas and Bronars 1989; Yinger 1998). Shifting the locus of prejudice from managers to consumers does not obviously increase modeling costs or reduce theoretical virtues such as simplicity and fruitfulness. And in agreement with Competitive Evaluation, we have supplied an alternative model.

3.3.2 Statistical discrimination

When Arrow (1972; 1973) objected that taste-based discrimination does not always allow hiring discrimination to persist under perfect competition, he did not use this insight to argue against the prevalence of hiring discrimination. Rather, Arrow developed a new model of what has come to be known as statistical discrimination (Fang and Moro

2010). Here is a simple presentation of statistical discrimination loosely following (Phelps 1972a,b).

Models of statistical discrimination remove the assumption that all workers are equally productive. Instead, the productivity f_j of members of each group j is modeled as normally distributed with $f_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Interviews provide a noisy signal $\overline{f_j}$ of workers' true productivity, with $\overline{f_j} = f_j + \epsilon_j$ for normally distributed noise $\epsilon_j \sim \mathcal{N}(0, \sigma_{\epsilon_j}^2)$. Noise ϵ_j is independent of true productivity f_j.

This model allows discrimination in the sense that two workers with equal quality signals $\overline{f}$ can be assigned different expected productivities by the employer, leading one to be hired and the other not hired based on identical interview performance. To see how this arises, note that expected productivity depends on both the signal $\overline{f}$ and the group mean μ_j as:

$$E[(f_j|\overline{f_j})] = \frac{\sigma_j^2 \overline{f_j} + \sigma_{\epsilon_j}^2 \mu_j}{\sigma_j^2 + \sigma_{\epsilon_j}^2} \tag{4}$$

Discrimination arises in two cases.

First, the noisiness $\sigma_{\epsilon_j}^2$ of productivity signals may vary across groups, for example because members of group A find it harder to get to know members of group B. As noise $\sigma_{\epsilon_j}^2$ increases, posterior expected productivity $E[(f_j|\overline{f_j})]$ tends towards the group mean μ_j, making it difficult for members of group B to distinguish themselves in competitive job markets.

Second, the group mean productivity μ_j may vary, for example because of historical discrimination (Coate and Loury 1993; Lundberg and Startz 1983) or in more recent models, because the actual distribution f_j of productivity may be replaced with an employer's prejudiced and inaccurate beliefs about productivity (Bohren et al. 2025). Lower group means will systematically reduce evaluations of group members, with the effect again increasing as evaluation becomes noisier.

All of these effects can persist in perfectly competitive equilibrium, and even become

self-reinforcing as B workers find themselves disincentivized to make costly human capital investments that are unlikely to be recognized by employers (Coate and Loury 1993; Lundberg and Startz 1983).

In agreement with the F-Twist, modeling statistical discrimination improves predictive accuracy in many domains (Fang and Moro 2010). The model above is neither complex nor costly to apply. The ongoing popularity of statistical discrimination models suggests that they are fruitful. And in agreement with Competitive Evaluation, we have provided a competing model.

3.3.3 Deviations from perfect competition

While the previous models aim to reconcile hiring discrimination with perfect competition, a more direct response would be that the existence of hiring discrimination reveals descriptively important ways in which markets often fail to be perfectly competitive. We briefly considered one response of this type in Section 3.3.2, where we saw that some recent statistical discrimination models allow employers to have prejudicial and inaccurate beliefs about group means, violating the assumption of perfect information. Beyond this, most markets in the 1950s exhibited at least one sizable barrier to entry: discrimination in credit markets preventing the entry of less-discriminatory minority-owned businesses (Pager and Shepherd 2008). Minority-owned businesses cannot outcompete discriminatory firms if they cannot secure a bank loan. Other popular models allow hiring discrimination to persist due to factors such as implicit bias (Bertrand et al. 2005), emotions such as fear and disgust (Rathelot and Safi 2023), and the costs of employment search (Black 1995).

Again, in agreement with the F-Twist classic presentations of these models have defended their predictive power. The growing popularity of these models suggests that they are fruitful and neither unduly complex nor unduly costly. And in agreement with Competitive Evaluation, each provides a competing model.

3.4 Taking stock

This section reviewed Milton Friedman's (1953) defense of idealization in "The methodology of positive economics." We saw that Friedman begins by offering classic arguments for idealization, but then infers based on very little argument to the wide applicability of perfect competition models and to skepticism about the extent of racial discrimination in hiring (Section 3.1). We saw that although some forms of hiring discrimination may be eliminated by perfect competition (Section 3.2), many forms of hiring discrimination persist under perfect competition and still others reveal important deviations from perfect competition (Section 3.3). In agreement with the Misapplication Thesis, we found that Friedman's own arguments for idealization do not support his model over competitors. This suggests that the disagreement about hiring discrimination between Friedman and critics need not turn on views about permissible descriptive idealization, but rather on which descriptive idealizations are justified by these views.

4 Greco on belief and credence

What is the relationship between belief and credence? Following Elizabeth Jackson (2020), let *belief-eliminativism* be the view that only credences, but not beliefs exist. Many theorists have challenged belief-eliminativism on the grounds that bounded agents must sometimes rely on full beliefs to simplify their cognitive lives. Section 4.1 considers five ways in which belief has been claimed to simplify the cognitive lives of bounded agents.

Daniel Greco's (2023) *Idealization in epistemology: A modest modeling approach* is one of the most prominent recent defenses of idealization in epistemological modeling. We will see in Section 4.2 that Greco takes his modest modeling approach to support belief-eliminativism as an idealizing assumption that can do all of the work which full beliefs were introduced to do. However, we will see in Section 4.3 that Greco's arguments fall short of establishing this conclusion. In agreement with the Misapplication Thesis, Greco's arguments fall short not because of any general opposition to idealization or challenges

to modest modeling, but simply because Greco does not show how credences play many of the simplifying roles of belief.

4.1 Simplifying roles of belief

Beliefs have been claimed to play at least five simplifying roles in the cognitive lives of bounded agents.

4.1.1 Beliefs simplify reasoning

One popular proposal is that beliefs simplify reasoning (Ross and Schroeder 2014; Staffel 2019):

> **(Beliefs Simplify Reasoning)** Reasoning with beliefs rather than credences often simplifies an agent's reasoning process.

Beliefs may simplify reasoning in at least two ways: by reducing the number of factors to consider, and by reducing the combinatorial complexity of considering them.

For example, suppose I want to decide whether to pack an umbrella today. Using full beliefs, I might reason:

(P1) It will rain today.

(P2) If it will rain today, then I should pack an umbrella.

∴ (C) I should pack an umbrella.

By contrast, using credences I might consider the possible world states $\Omega = \{w_i\}$, such as rain, snow and sunshine, alongside the possible acts $\mathcal{A} = \{a_j\}$ I could take. I could consider the consequences $a_j(w_i)$ of each act in each world state, assigning probabilities $Pr(w_i)$ to each state and utilities $u(a_j(w_i))$ to each consequence. I might then select the best act a^* as:

$$a^* = \underset{a_j \in \mathcal{A}}{\operatorname{argmax}} \sum_{w_i \in \Omega} Pr(w_i) u(a_j(w_i)). \tag{5}$$

This calculation requires me to consider many things: $|\Omega|$ states, $|\mathcal{A}|$ acts, $|\Omega| * |\mathcal{A}|$ consequences, $|\Omega|$ probabilities and $|\Omega| * |\mathcal{A}|$ utilities. And it requires a moderately complex calculation instead of a simple inference by modus ponens.

4.1.2 Beliefs simplify updating

A second view is that beliefs simplify updating (Harman 1986; Staffel 2019):

> **(Beliefs Simplify Updating)** Updating a body of beliefs is often simpler than updating a corresponding body of credences.

Beliefs may simplify updating by reducing both the frequency with which updates must be performed and the number of attitudes to be revised during updating.

For example, let p be the proposition that it will rain today. A minute from now, I have gained the evidence E that a minute has passed. On a credal view, I must update my credence in any proposition q for which $Pr(q|E) \neq Pr(q)$. This includes many propositions such as p, since E tells to a small degree against p. And this update must be performed often, since the passage of even a small interval of time bears on the likelihood of many related propositions, including p.

By contrast, beliefs towards p need not be updated so frequently. At most, they must be revised if the evidence is strong enough to shift which of belief, disbelief and suspension of judgment is now the appropriate attitude towards p. In Section 4.1.3 we will consider some reasons why updating may not be required even then. Moreover, updating may not involve changes in my attitudes towards a vast number of other propositions. For example, on one proposal learning that $\neg p$ should lead me to abandon a single belief, namely the belief that p, then to make minimal revisions to my other beliefs to restore consistency (Alchourrón et al. 1985). These minimal changes need not include changing my coarse-grained attitudes towards every proposition q to which $\neg p$ is probabilistically relevant.

4.1.3 Beliefs close inquiry

A third view is that agents must draw some inquiries to a close in order to limit resource expenditures (Thorstad 2024b) and simplify their research agenda (Olsson and Westlund 2006). Many theorists propose that beliefs close inquiry. Focusing for simplicity on the special case of inquiries into questions of the form 'whether p', the proposal is:

> **(Beliefs Close Inquiry)** Adopting a belief that p or a belief that $\neg p$ closes inquiry into the question 'whether p'.

On one account inspired by Michael Bratman's (1987) work on intention, beliefs close inquiry because they involve a commitment to bring inquiry to a close and to resist reopening inquiry without good reason (Holton 2014). On another account, inquiry characteristically involves a class of interrogative attitudes, all of which involve suspension of judgment (Friedman 2017). Adopting a belief that p or that $\neg p$ involves no longer suspending judgment on whether p, and hence closes inquiry into the question of whether p.

By contrast, the leading credal proposal for how inquiry closes is that inquiry closes when the expected value of future inquiry falls beneath its expected costs (Howard 1966). This proposal is sensitive to a more complex web of factors including the inquiries that might be performed, the results that they might yield, and the values and costs of these results. Moreover, this proposal must be further expanded to explain how and why agents resist reopening inquiries once they are closed. The resulting story may be still more complex and computationally demanding.

4.1.4 Beliefs coordinate complex reasoning

Sometimes, humans engage in complex reasoning processes involving many steps and even many component sub-inferences. It is not always possible to keep all steps of complex reasoning in view at once, creating a need to ensure that all of the various steps of complex reasoning processes are suitably coordinated. Some authors propose (Holton 2014):

(Beliefs Coordinate Complex Reasoning) Agents coordinate complex reasoning processes by using beliefs as fixed points in reasoning.

In more detail, the idea is that beliefs are more stable than typical credences are (Holton 2014; Staffel 2025). This happens because more evidence may be required for belief revision, and because beliefs may come coupled with a commitment to resist re-opening deliberation. Because beliefs are relatively stable, they can be taken as premises in complex reasoning without significant fear that premise beliefs may change during complex reasoning processes.

By contrast, credences may be less stable, leading to a live concern that the credences used in early steps of reasoning may no longer be in line with those used in later steps. Coordinating complex reasoning using credences may then require revisiting early reasoning steps, and this in turn could create a need to revisit later reasoning steps, substantially complicating the process of coordinating complex reasoning.

4.1.5 Beliefs facilitate active unopinionation

One way that humans simplify our cognitive lives is by remaining unopinionated. We are *passively unopinionated* about some propositions p towards which we lack doxastic attitudes of any type. But we are also *actively unopinionated* about some propositions which we have considered and formed doxastic attitudes about, because we take ourselves not to be in a position to form any more opinionated doxastic attitude. What does active unopinionation amount to?

A popular proposal is:

(Active Unopinionation as Suspension of Judgment) Agents are actively unopinionated towards a proposition if and only if they suspend judgment on that proposition.

Active Unopinionation as Suspension of Judgment fits naturally into a full-belief framework, where suspension of judgment has long been regarded as the counterpart to belief

and disbelief. Such frameworks admit many theories of the nature of suspension (Friedman 2017; McGrath 2021), many of which productively connect to other theoretical roles of belief. For example, on one story introduced in Section 4.1.3, agents close inquiry by moving from a state of suspension to a state of belief.

Credal accounts of active unopinionation model active unopinionation using states such as middling credence (Hájek 1998) or broad sets of imprecise credences (Van Fraassen 1989). However, not all of these accounts are naturally understood as forms of unopinionation (Friedman 2013). To say, for example, that I assign 50% credence to a fair coin landing heads need not imply that I am unopinionated about the outcome of the coinflip. Other accounts may not play the simplifying role of unopinionation in agents' cognitive lives. For example, many decision theories for imprecise credence are more, rather than less computationally demanding than counterpart theories for precise credence (Troffaes 2007).

4.1.6 Taking stock

So far, we have seen that beliefs have been claimed to play at least five simplifying roles in the lives of cognitively limited agents. Beliefs simplify reasoning (Section 4.1.1) and updating (Section 4.1.2), close inquiry (Section 4.1.3), coordinate complex reasoning (Section 4.1.4) and facilitate active unopinionation (Section 4.1.5).

The next order of business is to consider Greco's (2023) strategy for taking credence to play the simplifying role ascribed to full belief (Section 4.2) and to see how this strategy performs on the simplifying roles surveyed here (Section 4.3).

4.2 Modest modeling and belief-eliminativism

Greco (2023) rejects the *ambitious modeling* approach of seeking a complete and perfectly accurate theory of a domain. In its place, Greco proposes *modest modeling*. In selecting modeling frameworks, Greco advocates:

> **(Modesty in Framework Selection)** A modest modeler may use different frameworks to model different phenomena, relying essentially on good judgment in making her selection, and without any hope of an uber-framework that would provide systematic rules for which framework to use when. (Greco 2023, p. 36)

And in building models within a given framework, Greco advocates:

> **(Modesty in Model Selection)** Once the modest modeler has settled on a framework to use to model some phenomenon, the question of which model within that framework to construct may be similarly unsystematic. Perhaps the different models a framework allows the modeler to construct are in principle incapable of being merged into a single model, and good judgment is required to determine which models to use for which purposes. (Greco 2023, p. 36)

Greco uses his modest modeling approach to defend a variety of philosophical models and modeling frameworks that have been criticized for being overly idealized.

As Greco notes, modest modeling is compatible with the use of distinct and even incompatible models within a same domain. Nevertheless, Greco defends belief-eliminativism by arguing that all roles purportedly served by beliefs can be at least as well served by credences.

Greco considers the argument that full beliefs are needed to play a simplifying role in the cognitive life of bounded agents. Greco responds that the same simplifying role can be played by 'small world' Bayesian decision problems. Greco's main argument is brief, and is worth quoting in full:

> Rather than thinking of agents as always solving computationally demanding "grand-world" choices—decision problems in which the outcome space always includes all conceivable ways the rest of world history might progress, depending on all conceivable ways the agent might move their body at the

> time of choice—the modest Bayesian can think of them as solving much easier "small-world" choices—e.g., which of various breakfast cereals to eat, whether to call or fold in some game of poker. The Bayesian modeler can agree with the writers just quoted that limited agents like us need to make a wide range of non-trivial assumptions—about what our options are, what outcomes they might lead to, and how good or bad those outcomes would be—to tractably reason about what to do. She can just insist that those assumptions are properly reflected in the sorts of small-world Bayesian models we can actually write down and solve. Folk psychological notions of belief—understood as something over-and-above or distinct from the fact that small-world models have non-trivial state spaces Ω, in which plenty of logical possibilities are not represented—needn't come into the picture at all. (Greco 2023, p. 104)

Greco briefly considers and rejects the demand for further clarification of how small-world decision problems are formed and solved, then concludes that small-world decision problems allow credences to play the simplifying role that opponents have reserved for full belief.

4.3 Revisiting the belief roles

By all appearances, the disagreement does not turn on Criteria for Permissible Idealization, but rather on the specific question of whether credences can play the simplifying roles attributed to full belief. This question can be asked within or outside of a modest modeling perspective.

We will see below that the bare turn to small-world decision problems allows credences to play some of the simplifying roles attributed to full belief. But unless more is said, many roles will not be played. In agreement with the Misattribution Thesis, this suggests that Greco has taken a particular set of criteria for permissible idealization, namely the modest modeling perspective, to support idealizations which the perspective has not been shown

to support.

4.3.1 Simplifying reasoning

The first simplifying role for beliefs was:

> **(Beliefs Simplify Reasoning)** Reasoning with beliefs rather than credences often simplifies an agent's reasoning process.

We saw that beliefs simplify reasoning by reducing the number of factors to be considered as well as the combinatorial complexity of considering them.

The turn to small worlds goes same way towards reducing the number of factors to be considered, though not always as far as might be hoped. It does not directly address the computational complexity of the calculations to be made on the basis of those factors. To see these points, return to our example of deciding whether to pack an umbrella.

A small-world decision model of this situation might consider two acts: bringing and leaving an umbrella. It would consider two world-states: rain and non-rain. It would assign probabilities to each state and consequences to each act in each state, a total of four consequences. It would assign utilities to each of the four consequences. This is a non-trivial reduction in the number of factors to be considered, but still considerably more than those involved in our modus ponens inference. And the decision rule is unchanged. Agents must still select the act a^* maximizing

$$a^* = \underset{a_j \in \mathcal{A}}{\operatorname{argmax}} \sum_{w_i \in \Omega} Pr(w_i) u(a_j(w_i)). \tag{6}$$

The reduction in factors reduces the complexity of applying this decision rule, but on many views the complexity will still be beyond that of a simple modus ponens inference.

4.3.2 Simplifying updating

The second simplifying role for beliefs was:

(Beliefs Simplify Updating) Updating a body of beliefs is often simpler than updating a corresponding body of credences.

We saw that beliefs simplify updating by reducing both the frequency with which updates must be performed and the number of attitudes to be revised during updating.

Small world decision problems are generally regarded as models of how agents approach a single decision, rather than as models of the agent's stored beliefs. As such, the bare turn to small world decision problems should do little to simplify updating. Greco may mean to apply some sort of small-world assumption to an agent's stored beliefs. But some care is needed here, since agents require a great number of stored beliefs in order to guide thought and action. There is only so much simplification that can be done here, and conditionalizing the resulting body of beliefs is not likely to be an easy task.

4.3.3 Closing inquiry

A third simplifying role for beliefs was:

(Beliefs Close Inquiry) Adopting a belief that p or a belief that $\neg p$ closes inquiry into the question 'whether p'.

Greco does not propose any way in which credences close inquiry. If credences are meant to play this simplifying role, the argument must lie elsewhere.

4.3.4 Coordinating complex reasoning

A fourth simplifying role for beliefs was:

(Beliefs Coordinate Complex Reasoning) Agents coordinate complex reasoning processes by using beliefs as fixed points in reasoning.

Because beliefs are relatively stable, they can be taken as premises in complex reasoning without significant fear that premise beliefs will change during complex reasoning

processes. By contrast, credences are less stable and therefore increase the difficulty of coordinating complex reasoning processes which use them.

Greco does not propose any way in which the turn to small world decision problems may help credences to coordinate complex reasoning. It is certainly true that reasoning within small-world decision problems goes some way towards increasing stability by decreasing the number of propositions towards which agents might change their attitudes during reasoning. But fixing a given body of propositions, we have not yet been told why credences should be as stable as beliefs during complex reasoning on this body of propositions, nor given any other way in which credences better coordinate complex reasoning processes.

4.3.5 Facilitating active unopinionation

A final role for belief was facilitating active unopinionation. On a common view:

> **(Active Unopinionation as Suspension of Judgment)** Agents are actively unopinionated towards a proposition if and only if they suspend judgment on that proposition.

We saw that accounting for suspension of judgment as a form of active opinionation may be easier in models which assume full beliefs rather than, or alongside credences.

Greco does not propose any way in which the turn to small world decision problems should help credences to capture active unopinionation. If there is an argument to be made here, it lies elsewhere.

4.4 Taking stock

We saw in Section 4.1 that full belief has been claimed to play at least five important simplifying roles in the cognitive life of bounded agents. Full belief simplifies reasoning and updating, closes inquiry, coordinates complex reasoning, and facilitates active unopinionation. We saw in Section 4.2 that Greco (2023) argues for belief-eliminativism by

holding that the turn to small-world decision problems shows how credences can play the simplifying roles normally ascribed to full belief. However, we saw in Section 4.3 that Greco has not established this point. We saw that the bare turn to small-world decision problems goes some way towards simplifying reasoning, though not perhaps as far as we would like. We saw that the bare turn to small-world decision problems does not directly yield significant improvement on the other four roles.

If this is right, then what appears to be a debate between ambitious and modest approaches to modeling may be more productively understood through the Misapplication Thesis. Modest modelers argue for belief-eliminativism by arguing that credences play all, or nearly all of the roles ascribed to belief while also having theoretical advantages. But that is not what they have established. The problem lies not with modest modeling, as in a debate about Criteria for Permissible Idealization. The problem is rather a failure to show how modest modeling supports belief-eliminativism, as the Misapplication Thesis suggests.

5 Discussion

We saw in Section 1 that debates about idealization and bounded rationality often seem to turn on:

> **(Degree of Permissible Idealization)** For any subject S and purpose P, how idealized should models of S for purpose P be?

By contrast, I suggested that many debates turn neither on Degree of Permissible Idealization nor even on Criteria for Permissible Idealization. Bounded rationality theorists can often appeal instead to:

> **(Misapplication Thesis)** For some subject S, purpose P and descriptive idealization I involving cognitive bounds, I does not meet, or has not been shown to meet the criteria for permissible idealization accepted by the theorist proposing I.

Section 2 offered preliminary arguments for thinking that the Misapplication Thesis is often at issue: the Argument from Bounded Idealization and the Argument from Ideal Variation. Sections 3-4 showed how the Misapplication Thesis gets a take on two leading defenses of idealization. In both cases, what looks like a dispute about Degree of Permissible Idealization is in fact a failure to show that idealizations away from limited information, racial bias, or full belief are justified by shared standards for permissible idealization.

This discussion has important consequences for the distinction between ideal and non-ideal epistemology (Section 5.1), the justificatory burdens of idealization and de-idealization (Section 5.2), and the prospects for Bayesian approaches to bounded rationality (Section 5.3).

5.1 Beyond ideal and non-ideal epistemology

Sometimes the key to making progress is to ensure that debates are correctly framed. The debate between ideal and non-ideal epistemologists is often framed as a debate about the degree of, or criteria for permissible idealization. This framing is mirrored in the terminology used to describe the opposing positions. However, we saw in this paper that idealization as such is often not the main locus of disagreement.

More generally, Greco (2023) has argued that there could be no feasible project of de-idealizing epistemology. Idealization is a methodological tool to be used when appropriate, not opposed on principle. To stake out a claim against idealization is to stake out a peculiar perspective that is often removed from the specific considerations that bounded rationality theorists wish to incorporate into epistemological models.

Political philosophers have grown increasingly frustrated with the dichotomy between ideal and non-ideal theory. There are growing calls to abandon the dichotomy and reframe debates in other terms (McPherson and Plunkett 2025; Wiens 2025). Epistemologists might do well to make the same call.

What might replace the dichotomy between ideal and non-ideal epistemology? Many

distinctions have been brought under this dichotomy (Wiens 2025), which is part of why calls for abandonment are growing. But one central question at the heart of bounded rationality theorizing is the appropriateness of descriptive idealizations involving cognitive bounds. Bounded rationality theorists push for more such idealizations, and unbounded rationality theorists push for fewer.

Humans have limited cognitive abilities and must ration scarce cognitive resources such as attention (Castro and Pham 2020; Watzl 2017), memory (Harman 1986; Friedman 2018) and computational power (Icard 2018; Morton 2017). As modelers, we cannot always model each dimension of scarcity, nor can we ever fully capture any given dimension. But despite growing epistemological attention to cognitive bounds, cognitive bounds are often absent from epistemological discussions to which they are relevant. Bounded rationality theorists think that in many situations, this absence is worth addressing.

We saw in Section 2 that in political philosophy, even self-styled ideal theorists appeal to cognitive bounds to ground key claims. As such, there may be relatively broad appeal to the claim that incorporating cognitive bounds could strengthen models in some domains. And we may be able to make progress in stalled debates by focusing on more specific questions about when, why and in what way bounds should be incorporated.

5.2 The burdens of idealization

Many philosophers complain that idealized models are claimed from the armchair to explain any number of messy realities, while insufficient work is done to flesh out the relevant explanation. In this vein, we might understand Charles Mill's (2005) contention that Rawlsian theorists provide few useful insights about justice in societies marked by systemic racism, and indeed that there are many things about racist societies that cannot be understood without incorporating systemic racism at the outset of theorizing. We might also include Robin McKenna's (2023) suggestion that individualistic epistemological theories shed little light on important social realities, or Kristie Dotson's (2019) suggestion that epistemological theories are ill-equipped to understand the reasons for black rage against

state-sanctioned violence. We saw in the case studies that such complaints sometimes have merit. Friedman and Greco provide brief arguments for modeling assumptions which fall well short of establishing the claimed explanatory applications to hiring discrimination and belief-eliminativism.

Beyond their explanatory reach, we saw in Sections 3-4 that models are often justified against a range of further criteria. Criteria for permissible idealization impose a burden to show how proposed idealizations meet the relevant criteria. This includes saying in detail how the model accomplishes the explanatory task it is put forth to do, and how it realizes additional virtues such as simplicity and computational tractability. And as Friedman stresses, this discussion should make clear why the same benefits are unlikely to be realized by models which remove the proposed idealization. One way to read the lesson of this paper is that proponents of idealized models sometimes err by failing to meet the burden of showing in a competitive way why their models perform best on shared criteria for permissible idealization.

At the same time, criteria for permissible idealization impose a parallel burden on proponents of de-idealized models to show why removing idealizations leads to models which perform better on virtues such as explanatory power, simplicity and computational tractability. If Greco is right that there is no defensible project of de-idealization for its own sake, then the removal of idealizing assumptions is no less in need of defense than their inclusion, particularly since the removal of some idealizing assumptions is often coupled with the addition of others. In this way, the Misapplication Thesis might be productively understood not simply as the negative project of arguing against idealized models, but also as part of the positive project through which bounded rationality theorists meet the burden of competitively justifying their models.

5.3 Bayes and bounds

Many critics of bounded rationality have suggested that the tradition makes little room for Bayesian theorizing. Robert Pasnau (2013) suggests that Bayesian models are so

idealized that they could only ever belong to ideal epistemology. Likewise, Jennifer Carr (2022) characterizes non-ideal epistemology as informal, non-mathsy, written in Word, emphasizing knowledge and concerned with unimpressive babies, whereas ideal epistemology is characterized as formal, mathsy, written in LaTeX, emphasizing decision theory and concerned with superbabies.

Recent years have seen a rapid growth in the number and success of Bayesian approaches to bounded rationality. Bounds have been incorporated into key paradigms such as resource-rational analysis (Icard ms; Lieder and Griffiths 2020), boundedly rational analysis (Icard 2018), and cognitively bounded rational analysis (Howes et al. 2009). Contemporary Bayesian models involve agents with limited attention (Maćkowiak et al. 2023), memory (Anderson 1990) and computational power (Lieder and Griffiths 2020), who update occasionally (Dallman 2017), reason heuristically (Chater and Oaksford 1999a) and consider only small samples of stored information (Icard 2018; Vul et al. 2014). This trend towards incorporating bounds into Bayesian models and paradigms suggests that there may be a place for Bayesian theorizing within the bounded tradition.

We can make sense of this trend if we view bounded rationality not as a blanket opposition to idealization as such, but rather as an insistence on the importance of modeling cognitive bounds in key situations of interest. From this perspective, the mistake made by traditional Bayesian models is not the use of idealization, mathematics, probability theory, decision theory or LaTeX. It is rather the use of these tools in a way that fails to incorporate relevant cognitive bounds. When cognitive bounds are built into Bayesian models in relevant, tractable and explanatorily powerful ways, these models gain the capacity to meaningfully illuminate norms and behaviors in a wide variety of situations involving bounded agents. Such models are welcome contributions to the theory of bounded rationality.

References

Alchourrón, Carlos, Gärdenfors, Peter, and Makinson, David. 1985. "On the logic of theory change: Partial meet contraction and revision functions." *Journal of Symbolic Logic* 50:510–530.

Anderson, John. 1990. *The adaptive character of thought*. Lawrence Erlbaum Associates.

Arrow, Kenneth. 1972. "Some mathematical models of race in the labor market." In Anthony Pascal (ed.), *Racial discrimination in economic life*. Lexington Books.

—. 1973. "The theory of discrimination." In Orley Ashenfelter and Albert Rees (eds.), *Discrimination in labor markets*, 3–33. Princeton University Press.

Becker, Gary. 1957. *The economics of discrimination*. University of Chicago Press.

Bertrand, Marianne, Chugh, Dolly, and Mullainathan, Sendhil. 2005. "Implicit discrimination." *American Economic Review* 95:94–8.

Black, Dan. 1995. "Discrimination in an equilibrium search model." *Journal of Labor Economics* 13:309–34.

Bohren, J. Aislinn, Haggag, Kareem, Imas, Alex, and Pope, Devin. 2025. "Inaccurate statistical discrimination: An identification problem." *Review of Economics and Statistics* 107:605–20.

Borjas, George and Bronars, Stephen. 1989. "Consumer discrimination and self-employment." *Journal of Political Economy* 97:581–605.

Bowers, Jeffrey and Davis, Colin. 2012. "Bayesian just-so stories in psychology and neuroscience." *Psychological Bulletin* 138:389–414.

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

Carr, Jennifer. 2022. "Why ideal epistemology?" *Mind* 131:1131–62.

Cartwright, Nancy. 1989. *Nature's capacities and their measurement*. Oxford University Press.

Castro, Clinton and Pham, Adam. 2020. "Is the attention economy noxious?" *Philosophers' Imprint* 20:1–13.

Chamberlin, Edward. 1923. *The theory of monopolistic competition*. Harvard University Press.

Chater, Nick and Oaksford, Mike. 1999a. "The probability heuristics model of syllogistic reasoning." *Cognitive Psychology* 258:191–258.

—. 1999b. "Ten years of the rational analysis of cognition." *Trends in Cognitive Sciences* 3:57–65.

Coate, Stephen and Loury, Glenn. 1993. "Antidiscrimination enforcement and the problem of patronization." *American Economic Review* 83:92–8.

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

Dawes, Robyn. 1979. "The robust beauty of improper linear models in decision making." *American Psychologist* 34:571–582.

Dawes, Robyn and Corrigan, Bernard. 1974. "Linear models in decision making." *Psychological Bulletin* 81:95–106.

Dotson, Kristie. 2019. "Tales from an apostate." *Philosophical Issues* 29:69–83.

Elliott-Graves, Alkistis and Weisberg, Michael. 2014. "Idealization." *Philosophy Compass* 9:176–85.

Fang, Hanming and Moro, Andrea. 2010. "Theories of statistical discrimination and affirmative action: A survey." In Jess Benhabib, Alberto Bisin, and Matthew Jackson (eds.), *Handbook of Social Economics*, 133–200. North Holland.

Frazer, William and Boland, Lawrence. 1983. "An essay on the foundations of Friedman's methodology." *American Economic Review* 73:129–44.

Friedman, Jane. 2013. "Rational agnosticism and degrees of belief." *Oxford Studies in Epistemology* 4:57–81.

—. 2017. "Why suspend judging?" *Noûs* 51:302–326.

—. 2018. "Junk beliefs and interest-driven epistemology." *Philosophy and Phenomenological Research* 97:568–83.

Friedman, Milton. 1953. "The methodology of positive economics." In *Essays in positive economics*. University of Chicago Press.

Gigerenzer, Gerd and Selten, Reinhard (eds.). 2001. *Bounded rationality: The adaptive toolbox*. MIT Press.

Greco, Daniel. 2023. *Idealization in epistemology: A modest modeling approach*. Oxford University Press.

Hájek, Alan. 1998. "Agnosticism meets Bayesianism." *Analysis* 58:199–206.

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

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.

Holton, Richard. 2014. "Intention as a model for belief." In Manuel Vargas and Gideon Yaffe (eds.), *Rational and social agency: Essays on the philosophy of Michael Bratman*, 12–37. Oxford University Press.

Howard, Ronald. 1966. "Information value theory." *IEEE Transactions on Systems Science and Cybernetics* 2:22–6.

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.

Hughes, Nick. 2024. "Non-ideal epistemic rationality." *Philosophical Issues* 34:72–95.

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

—. ms. *Resource rationality*.

Jackson, Elizabeth. 2020. "The relationship between belief and credence." *Philosophy Compass* 15:e12668.

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

Jones, Matt and Love, Bradley. 2011. "Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition." *Behavioral and Brain Sciences* 34:169–231.

Klaes, Matthias and Sent, Esther-Mirjam. 2005. "A conceptual history of the emergence of bounded rationality." *History of Political Economy* 37:27–59.

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.

Lundberg, Shelly and Startz, Richard. 1983. "Private discrimination and social intervention in competitive labor markets." *American Economic Review* 73:340–7.

Maćkowiak, Bartosz, Matějka, Filip, and Wiederholt, Mirko. 2023. "Rational inattention: A review." *Journal of Economic Literature* 61:226–73.

Mäki, Uskali (ed.). 2009. *The methodology of positive economics: Reflections on the Milton Friedman legacy*. Cambridge University Press.

Martignon, Laura. 2001. "Comparing fast and frugal heuristics and optimal models." In Gerd Gigerenzer and Reinhard Selten (eds.), *Bounded rationality: The adaptive toolbox*. MIT Press.

McCloskey, Deirdre. 1985. *The rhetoric of economics*. University of Wisconsin Press.

McGrath, Matthew. 2021. "Being neutral: Suspension of judgement, agnosticism and inquiry." *Noûs* 55:463–84.

McKenna, Robin. 2023. *Non-ideal epistemology*. Oxford University Press.

McMullin, Ernan. 1985. "Galilean idealization." *Studies in History and Philosophy of Science* 16:247–73.

McPherson, Tristram and Plunkett, David. 2025. "Conceptual ethics and the categories of 'ideal theory' and 'non-ideal theory' in political philosophy: A proposal for abandonment." In Manuel Gustavo Isaac, Steffen Koch, and Kevin Scharp (eds.), *New perspectives on conceptual engineering*, volume 2, 73–94. Springer.

Mills, Charles. 2005. "'Ideal theory' as ideology." *Hypatia* 20:165–84.

Morton, Jennifer. 2017. "Reasoning under scarcity." *Australasian Journal of Philosophy* 95:543–59.

Musgrave, Alan. 1981. "'Unreal assumptions' in economic theory: The F-twist untwisted." *Kyklos* 34:377–87.

Oaksford, Mike and Chater, Nick. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Olsson, Erik J. and Westlund, David. 2006. "On the role of the research agenda in epistemic change." *Erkenntnis* 65:165–83.

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

Pasnau, Robert. 2013. "Epistemology idealized." *Mind* 122:987–1021.

Phelps, Edmund. 1972a. *Inflation policy and unemployment theory: The cost-benefit approach to monetary planning*. Macmillan.

—. 1972b. "The statistical theory of racism and sexism." *American Economic Review* 62:659–61.

Rathelot, Roland and Safi, Mirna. 2023. "Taste-based discrimination." In Ashwini Deshpande (ed.), *Handbook on economics of discrimination and affirmative action*, 17–48. Springer.

Robinson, Joan. 1933. *The economics of imperfect competition*. Macmillan.

Ross, Jacob and Schroeder, Mark. 2014. "Belief, credence, and pragmatic encroachment." *Philosophy and Phenomenological Research* 88:259–288.

Schmidtz, David. 2011. "Nonideal theory: What it is and what it needs to be." *Ethics* 121:722–96.

Simmons, A. John. 2010. "Ideal and nonideal theory." *Philosophy and Public Affairs* 38:5–36.

Simon, Herbert. 1947. *Administrative behavior: a study of decision-making processes in administrative organization*. Macmillan.

Staffel, Julia. 2019. "How do beliefs simplify reasoning?" *Noûs* 53:937–62.

—. 2020. *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.

—. 2025. *Unfinished business*. Oxford University Press.

Thorstad, David. 2024a. "The accuracy-coherence tradeoff in cognition." *British Journal for the Philosophy of Science* 75:695–715.

—. 2024b. *Inquiry under bounds*. Oxford University Press.

—. 2025. "The complexity-coherence tradeoff in cognition." *Mind* 134:422–57.

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

Van Fraassen, Bas. 1989. *Laws and symmetry*. Oxford University Press.

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.

Watzl, Sebastian. 2017. *Structuring the mind: The nature of attention and how it shapes consciousness*. Oxford University Press.

Wheeler, Gregory. 2020. "Less is more for Bayesians, too." In Riccardo Viale (ed.), *Routledge handbook on bounded rationality*, 471–83. Routledge.

Wiens, David. 2025. "Beyond ideal theory: A framework for methodological debates." *Philosophy Compass* 20:e70046.

Wong, Stanley. 1973. "The 'F-Twist' and the methodology of Paul Samuelson." *American Economic Review* 63:312–25.

Yinger, John. 1998. "Evidence on discrimination in consumer markets." *Journal of Economic Perspectives* 12:23–40.