The accuracy-coherence tradeoff in cognition

Forthcoming in *British Journal for the Philosophy of Science*

Penultimate draft, please cite published version

Abstract

I argue that bounded agents face a systematic accuracy-coherence tradeoff in cognition. Agents must choose whether to structure their cognition in ways likely to promote coherence or accuracy. I illustrate the accuracy-coherence tradeoff by showing how it arises out of at least two component tradeoffs: a coherence-complexity tradeoff between coherence and cognitive complexity, and a coherence-variety tradeoff between coherence and strategic variety. These tradeoffs give rise to an accuracy-coherence tradeoff because privileging coherence over complexity or strategic variety often leads to a corresponding reduction in accuracy. I conclude with a discussion of two normative consequences for the study of bounded rationality: the importance of procedural rationality and the role of coherence in theories of bounded rationality.

1 Introduction

It is often held that cognitively unbounded agents are rationally required to satisfy a range of coherence requirements. For example, unbounded agents should have probabilistically coherent credences, logically consistent beliefs, and transitive preferences.

Theories of bounded rationality tend to relax the requirement of full coherence. This is done because bounded agents are often unable to achieve full coherence. Nevertheless, many theorists hold that there is a strong normative role for coherence in the theory of bounded rationality (Staffel 2020; Zynda 1996). For example, bounded agents might be rationally required to be as coherent as possible given their limitations.

In support of the normative importance of coherence, it has recently been shown how traditional arguments in favor of full coherence requirements on unbounded agents can be generalized to support approximate coherence requirements on bounded agents. These include arguments from accuracy-dominance (De Bona and Staffel 2017, 2018; Staffel 2020) and vulnerability to book (Schervish et al. 2000, 2002, 2003; Staffel 2015).

At the same time, some leading scientific approaches to bounded rationality place less emphasis on coherence (Arkes et al. 2016).[1] One such approach holds that bounded agents should use a toolbox of fast-and-frugal heuristics which make efficient and accurate judgments at the price of significant vulnerability to incoherence (Gigerenzer and Selten 2001). What could be said in favor of this looser role for coherence in the theory of bounded rationality?

A defining feature of bounded rationality is the existence of tradeoffs. Most famously, there is often an accuracy-effort tradeoff between cognitive strategies (Johnson and Payne 1985).[2] The cognitive strategies which produce the most accurate judgments or the best decisions tend to be among the most effortful strategies available. As a result, bounded agents need to select cognitive strategies which strike a good balance between accuracy and effortfulness.

In this paper, I argue that bounded agents face a comparably systematic accuracy-coherence tradeoff. The strategies which produce the most accurate judgments or the best decisions come apart from the most coherent strategies. If the existence of an accuracy-coherence tradeoff can be demonstrated, it will put some pressure against the relationship between bounded rationality and approximate coherence by showing that agents often pay a price in accuracy for gained coherence.

Here is the plan. Section 2 clarifies what it means to say that there is an accuracy-coherence tradeoff in cognition. Section 3 introduces the accuracy-coherence tradeoff using an extended example. Sections 4-5 illustrate two of the cognitive factors driving the

[1]For philosophical overviews of the ensuing debates, see Rysiew (2008) and Sturm (2012).

[2]Sometimes there is no accuracy-effort tradeoff. In some situations, the most accurate strategies are also among the most frugal (Geman et al. 1992; Gigerenzer and Brighton 2009; Wheeler forthcoming).

accuracy-effort tradeoff: cognitive complexity and strategic variety. Section 6 resolves a puzzle about the location of the accuracy-coherence tradeoff. Section 7 uses the resolution of this puzzle to draw lessons about the shape and extension of coherence-based normative theories.

2 Clarifying the target

What does it mean to say that bounded agents face an accuracy-coherence tradeoff? In this section, I offer five remarks to clarify my target, building on the analogy between the accuracy-effort and accuracy-coherence tradeoffs.

First, the term 'accuracy-effort tradeoff' is standardly used to denote two different phenomena. On the one hand, it denotes an *accuracy-effort tradeoff* in which the accuracy of our judgments, measured perhaps by a scoring rule, trades off against the effort of making judgments. On the other hand, it denotes a *quality-effort tradeoff* in which the quality of our decisions, measured perhaps by a utility function, trades off against the effort of making decisions. When clarity is of the essence, I break with tradition and use the term 'accuracy-effort tradeoff' to designate only the first tradeoff. But it is no accident that the accuracy-effort and quality-effort tradeoffs have been lumped together, since they arise in similar cases and for structurally similar reasons. When an umbrella term is needed, I hold with tradition in using the term 'accuracy-effort tradeoff' to denote both of these component tradeoffs.

Similarly, the term 'accuracy-coherence tradeoff' encompasses both an *accuracy-coherence tradeoff* in judgment and a *quality-coherence tradeoff* in decisionmaking. When clarity is required, I reserve the term 'accuracy-coherence tradeoff' for the first tradeoff only. But we will see that the accuracy-coherence and quality-coherence tradeoffs arise in similar cases and for structurally similar reasons. For this reason, I build on existing tradition and use the umbrella term 'accuracy-coherence tradeoff' to designate both component tradeoffs when the meaning of this phrase is clear from context.

Second, the accuracy-coherence tradeoff, like the accuracy-effort tradeoff, occurs at the level of cognitive strategies rather than the judgments and decisions that they produce. We might express the point expectationally.[3] To say that there is an accuracy-coherence tradeoff in cognition is to say that agents are often forced to choose between strategies with better impacts on the expected coherence or expected accuracy of their beliefs, respectively. When the processes in question directly issue in judgments, the expected accuracy and coherence of a process will be the expected accuracy and coherence of the agent's resulting doxastic state. I am primarily concerned with examples of this form. We could extend the terminology to talk about the expected accuracy and coherence of other processes such as evidence-gathering in terms of their expected downstream contributions to the accuracy or coherence of an agent's beliefs, although this will not be my primary concern. Similar glosses can be given for the quality-coherence tradeoff.

Third, the accuracy-coherence tradeoff, like the accuracy-effort tradeoff, occurs often but not always. In some situations, the accuracy-effort tradeoff is nonexistent, or even reversed: agents can improve the expected accuracy of their judgments by using less-effortful cognitive processes (Geman et al. 1992; Gigerenzer and Brighton 2009). For this reason, the most important task in studying the accuracy-effort tradeoff is to characterize the conditions under which it occurs. Similarly, I argue that the accuracy-coherence tradeoff often occurs, but not that it always occurs. The most important task is to identify factors driving the accuracy-coherence tradeoff in cognition, and I argue that both strategic variety (§4) and cognitive complexity (§5) fit the bill.

Fourth, the notion of coherence is notoriously fraught. For one thing, different theorists employ a patchwork of at most roughly coextensional terms to pick out the phenomenon of interest, including 'coherence', 'consistency', 'structural rationality', 'axiomatic rationality' (Gigerenzer forthcoming), and the 'standard picture' of rationality (Stein 1996). For another, there is disagreement about which patterns of attitudes should count as incoher-

[3] If a non-expectational gloss is desired, we could also express the point in terms of the dispositions of processes to produce accurate and coherent belief.

ent. And in the practical domain, it is disputed which objects may count as incoherent: choices, choice-dispositions, or preferences.

To cut through these debates, I give examples which can be phrased in terms of attitudinal coherence involving judgments or preferences, but also rephrased using other standard descriptions if desired. I focus on requirements widely agreed to characterize coherent judgment and preference, such as the probability axioms and the requirement that strict preferences be asymmetric. And I give a variety of examples, each drawing on a different requirement of practical or theoretical coherence.

Fifth, because my concern is with processes of judgment and decisionmaking, most of my examples have a diachronic flavor. This raises a natural objection: what seems to be attitudinal incoherence could instead reflect changes of attitudes across time. For example, if I choose X over Y at t_1 and Y over X at t_2, this might not show that I held the incoherent preferences $X > Y$ and $Y > X$ together. I may have changed my preferences due to learning, changing tastes, or transformative experience in the interim. To handle this challenge, I discuss natural attempts to redescribe my examples by invoking preference change and argue that these attempts fail.

Summing up, in its simplest form the accuracy-coherence tradeoff is an umbrella term for a tradeoff between the expected accuracy and coherence of the judgments resulting from a process of judgment formation, as well as a related tradeoff between the expected quality and coherence of decisions resulting from a process of decisionmaking. The accuracy-coherence tradeoff occurs often, but not always, and my project will be to illustrate two factors which may drive the accuracy-coherence tradeoff. When possible, I focus on paradigmatic examples of practical or theoretical incoherence, describing the examples in terms of attitudinal incoherence. And I argue against natural attempts to recast these examples as diachronically coherent attitudinal change. With these remarks in mind, I turn to an example designed to illustrate how the accuracy-coherence tradeoff arises (Section 3), before exploring two factors which drive the accuracy-coherence tradeoff (Sections 4-5).

3 An example: lexicographic and semilexicographic choice

To get a grip on the many ways in which accuracy and coherence trade off for bounded agents, it will help to begin with an example from decisionmaking, then extend this example to a structurally similar example in judgment formation. Suppose you are buying a car. You might decide using lexicographic choice (Fishburn 1974). You would order features of cars by their importance. Perhaps the most important feature is that it has an automatic transmission; next most important is price; then other features such as safety rating and comfort follow. You would compare the available cars by their most important feature, choosing the car which scores best on this feature. If several cars score just as well on this feature, for example because they all have an automatic transmission, you would compare cars along the second-most-important feature, price, continuing in this way until a decision was reached.

More formally, lexicographic choosers confront a choice set $O = \{o_1, \ldots, o_m\}$. They select some decision cues $c_1, \ldots, c_n$ such as transmission type, price and safety rating, rank-ordered by importance. Lexicographic choosers then estimate or retrieve from memory the values $f(o_i, c_j)$ of each option along each cue. For example, $f(o_3, c_2)$ is the price of the third car. For each cue c_i, they determine a value function V_i ranking the goodness of each value that c_i can take. For example, perhaps $V_1(x) = 1$ if x is 'automatic', and 0 otherwise. If some option o maximizes V_1, then o is chosen. Otherwise, the options maximizing V_1 are compared according to V_2, repeating until one option remains.

Lexicographic choice is quite a silly way to buy a car.[4] Our lexicographic chooser will always buy the cheapest automatic car unless two automatic cars are tied in price. A traditional and cognitively efficient way to improve upon lexicographic choice is semilexicographic choice (Tversky 1969).[5] Semilexicographic choice fixes for each i a small difference

[4]To be clear, the claim is that lexicographic choice is a silly way to *buy a car*, not that lexicographic choice is always a silly way to make decisions.

[5]The version of semilexicographic choice that I present here is commonly studied in psychology and judgment and decisionmaking fields. Some theorists have developed other notions of semilexicographic choice (Manzini and Mariotti 2012).

σ_i in value which will be ignored.[6] For example, in the simple case where each V_i beyond the first is the identity function, we might choose to ignore price differences under $1,000 and safety-differences no greater than one star. The turn to semilexicographic choice is widely held to yield improvements in decision quality. A semilexicographic chooser will not buy the cheapest automatic car if a comparably-priced automatic is significantly safer.

But semilexicographic choice is less coherent than lexicographic choice.[7] To see the problem, suppose that three automatic cars have the following costs and safety ratings.

	Car A	**Car B**	**Car C**
Cost (Thousands of dollars)	19	18.6	17.8
Safety Rating (Stars)	4	2.5	1

If given the pairwise choice between Car A and Car B, our semilexicographic chooser will pick Car A. Between Car B and Car C, she will choose Car B. And between Car C and Car A, she will choose car C. It is natural to interpret this result as a collection of intransitive preferences.

Lexicographic choice does not have this problem. The pairwise choices made by a lexicographic chooser are always transitive. In this way, the move from lexicographic to semilexicographic choice is an instance of the quality-coherence tradeoff. Going semilexicographic decreases the coherence of an agent's decisions, but increases their quality.

The same tradeoff can be found in cases of judgment formation rather than decision-making. Consider a binary judgment problem, such as judging which of two cities is larger. A popular heuristic for binary judgment is *take the best* (TTB) (Gigerenzer and Goldstein 1996). TTB instructs agents to identify a number of judgmental cues which are relevant to the size of a city, such as the presence of an airport or its status as a national capitol. Agents then retrieve or estimate values of each cue from memory, ordering cues

6 That is, option o is eliminated at stage i just in case for some o' we have $V_i(f(o', c_i)) > V_i(f(o, c_i)) + \sigma_i$.

7 The observation that semilexicographic choices can be intransitive suffices to show that semilexicographic choice is vulnerable to some forms of incoherence that lexicographic choice avoids. By contrast, any incoherence in lexicographic choice will also be an incoherence in semilexicographic choice, of which lexicographic choice is a special case.

by predictive validity.[8] Agents compare cities along each cue, beginning with the most valid cues, until they find one cue on which one city performs better than the other. They then halt deliberation with the judgment that this city is larger.

In many contexts, TTB is an excellent way to form accurate judgments at low cognitive cost. But TTB is precisely the judgmental analog of lexicographic choice, and for that reason it suffers from the same vulnerability: a city which performs well on the highest-ranking cue will always be judged larger than any other city, no matter how poorly it performs on lower-ranking cues. Plausible ways of fixing this vulnerability purchase increased accuracy at the expense of vulnerability to incoherence.[9]

One way to fix the problem would be to go semilexicographic, ignoring small differences along each judgmental cue. That would produce intransitive judgments of comparative size in exactly the same way as before.[10] Here is another bugfix. Number the cues as $c_1, \ldots, c_k$ in descending order of predictive validity. For each cue c_i, let the *tally* of c_i be 1 if c_i favors City 1 being larger, -1 if c_i favors City 2, and 0 otherwise. Going down the list of cues, the *running tally* at i will be the sum $t_1 + \cdots + t_i$ of tallies from 1 to i. In this language, TTB says to halt when the running tally hits ± 1 and output the judgment that the currently-favored city is larger. Generalizing, *take the n-best* (TTNB) would say to halt when the running tally hits $\pm n$ for some fixed n, and output the judgment that the currently-favored city is larger.[11] Increasing n even moderately would go a long way towards erasing the risk of favoring cities which perform abnormally well on the most

[8] The validity of cue c_i is the probability that it successfully discriminates, given that it discriminates at all. More formally, let $A > B$ hold when city A is larger than city B, and $A >_i B$ hold when A outperforms B on cue c_i. Then within a reference class C, the validity of c_i is $|\{(X, Y) \in C \times C : X > Y \wedge X >_i Y\}| \,/\, |\{(X, Y) \in C \times C : X >_i Y\}|$.

[9] Like lexicographic choice, TTB is immune to the type of intransitivity illustrated below. It should also be emphasized that in many contexts, TTB is already fairly accurate (Armstrong and Graefe 2012; Gigerenzer and Goldstein 1996; Martignon and Hoffrage 2002).

[10] A bit more precisely, the agent judges of cities A,B,C that A is larger than B, B is larger than C, and C is larger than A. Plausibly, most competent speakers also believe that comparative size is transitive, a belief that is logically inconsistent with the previous three judgments.

[11] We will also need a tiebreaker stipulation, for example: after all cues are exhausted, the city with the largest tally is judged larger, and no judgment is returned in the case of a tie.

important cues.[12]

But TTNB also produces intransitive judgments of comparative size. Consider an example with three cities and five cues:

	City A	**City B**	**City C**
$\mathbf{c_1}$	0	0	1
$\mathbf{c_2}$	0	2	1
$\mathbf{c_3}$	2	1	0
$\mathbf{c_4}$	2	1	0
$\mathbf{c_5}$	2	1	0

Here TT2B produces the judgments that City A is larger than City B; B is larger than C; and C is larger than A. Similar examples can be produced for $n > 2$. The point, as before, is that natural ways of improving the accuracy of TTB, such as going semilexicographic or generalizing to TTNB, create new opportunities for incoherence.

This example lends plausibility to the idea that there could be a more general tradeoff between accuracy and coherence in cognition. But one example does not demonstrate a tradeoff, nor does it provide any explanatory illumination. To show that there is a systematic accuracy-coherence tradeoff, I consider two intermediate factors which trade off against coherence: strategic variety and cognitive complexity. In each case, I argue, privileging coherence over the intermediate factor often comes at the expense of accuracy. The result of this discussion will be two ways in which an accuracy-coherence tradeoff can be produced.

[12]While a full study of the accuracy of TTNB is beyond the scope of this paper, one way to see that it is likely to be an accuracy improvement is to note that for n large, TTNB converges to equal-weighted linear choice, or *tallying*, which in many contexts is held to perform about as well as linear regression (Dawes and Corrigan 1974; Dawes 1979).

4 Coherence and strategic variety

Cognitive strategies can produce incoherence in one of two ways. First, there is the risk of *internal incoherence*: the tendency of a single strategy to produce judgments or decisions that fail to cohere with other judgments and decisions produced by the same strategy. Second, there is the risk of *external incoherence*: the tendency of a strategy to produce judgments or decisions that fail to cohere with judgments and decisions produced by other strategies that the agent employs.

Some agents avoid the risk of external incoherence by using only a single cognitive strategy throughout their lifetimes. For example, agents can avoid external incoherence in their credences by always updating through Bayesian conditionalization. But it is usually held that heuristic cognizers do and should employ a large toolbox of different heuristic strategies for judgment and decisionmaking (Gigerenzer and Selten 2001). Because no heuristic strategy performs well in all environments, it would be inappropriate for heuristic cognizers to use the same strategy to solve all problems. Rather, heuristic cognizers learn to switch flexibly between various strategies according to the demands of each situation (Marewski and Schooler 2011; Payne et al. 1988).

Now consider a heuristic cognizer deciding how many cognitive strategies she will employ. Plausibly, many such agents confront an accuracy-coherence tradeoff in setting the size of their toolbox. On the one hand, up to a point, increasing the size of her toolbox increases the expected accuracy of the agent's judgments and expected quality of her decisions. This happens because increasing the size of her toolbox means that the agent is more likely to have appropriate strategies available to confront any given situation. On the other hand, increasing the size of her toolbox may decrease the expected coherence of an agent's beliefs and preferences. Each strategy added to the toolbox increases the risk of external incoherence by creating the real possibility that different strategies will be applied to similar or identical problems throughout her lifetime. Because there is no guarantee that different strategies will give similar answers to similar problems, the agent may be led to make radically different judgments or decisions in similar or identical judgment and choice problems.

Cashing out this plausibility argument in full generality would be a harrowing task. First, we would need to assess the effects of strategic variety on the lifetime judgments and decisions made by an agent. This task risks empirical intractability. And second, we would need to measure the accuracy and coherence of the resulting judgments and decisions. This task risks overreliance on particularities of our favorite measure.[13] However, we can illustrate the way in which increasing strategic variety often generates a quality-coherence tradeoff by focusing on cases of repeated decisionmaking, then modify this discussion to illustrate an accuracy-coherence tradeoff.

Suppose that every week you are faced with the choice between two brands of cereal. Here are two ways you might decide which cereal to purchase. First, you might decide using *considered defaults*. On your first trip to the grocery store, you would choose a cereal using a cognitively demanding heuristic, for example by *tallying* (Dawes and Corrigan 1974) the positive features on which each cereal comes out better, then choosing the cereal with the most positive features. You could spend a good deal of time gathering and retrieving information from memory and from your environment to inform this choice. On subsequent trips, you would choose by default (Johnson and Goldstein 2003), picking the same cereal that you first selected. You would continue in this way until reassessment was triggered, for example because a set amount of time had elapsed or new information indicated that redeliberation could be called for. Then you would once again choose using a cognitively demanding heuristic, subsequently continuing to choose by default as before.

Second, you might use a *consistent strategy method*, redeliberating each time using a simple heuristic such as lexicographic choice. This method would require you to use a simpler heuristic to make your initial decision, in order to balance deliberation costs, but would allow you to reopen deliberation during each shopping trip in order to respond to

[13] See Staffel (2020) for an overview of divergences for measuring degrees of coherence.

new information.

The considered default method differs from the consistent strategy method in that it mixes two different decisionmaking strategies. This opens the door to incoherence. Suppose that on your first trip (t_0), you choose Mighty Muesli (X) over Good Granola (Y). Gradually, the makers of Mighty Muesli begin shrinking the size of boxes and increasing their price, and you notice these changes. You continue choosing by default, so that on your penultimate trip (t_{n-1}) you choose Mighty Muesli (X') over Good Granola (Y'). Here the prospects X, Y have been replaced with X', Y' to reflect changes in decision-relevant features such as size and price. But on your last trip (t_n) redeliberation is triggered, for example because sufficient time has elapsed, or because the store comes under new management and this change suggests that redeliberation may be in order. Although the price, quality and other features of Mighty Muesli (X') and Good Granola (Y') have not changed since last week, you now choose Good Granola over Mighty Muesli based on the changes in price and quality since t_0.

It is often held as a requirement of coherence that strict preferences be asymmetric:

(Asymmetry of Strict Preference) For all prospects X, Y: $X \nsucc Y$ or $Y \nsucc X$.

Your choice of Mighty Muesli at t_{n-1} reveals the weak preference $X' \succcurlyeq Y'$, and your choice of Good Granola at t_n reveals the weak preference $Y' \succcurlyeq X'$. This is not yet a case of symmetric strict preference. But we can modify the example to induce symmetric strict preferences. Suppose that at t_{n-1}, Good Granola was discounted by ten cents, and represent this discounted prospect as Y^*. Your choice at t_{n-1} then reveals the weak preference $X' \succcurlyeq Y^*$, and presumably you enjoy free money, giving $Y^* \succ Y'$. It follows that at t_{n-1} you have $X' \succ Y'$, on pain of another form of incoherence, namely intransitivity. And by a similar device we can induce $Y' \succ X'$ at t_n. In this way, considered default choice can lead to incoherence in the form of symmetric strict preferences.

Because this example is diachronic, it might be objected that diachronic symmetry of preference can reveal rationally permissible preference change rather than diachronic incoherence. But the problem is that standard signs of diachronically coherent preference change are not present in this example. On the one hand, we want to allow that agents may sometimes change their fundamental preferences over time, for example by acquiring a taste for granola or by undergoing transformative experiences which rewrite their personality. But this is not what happened here. The agent has not, we may stipulate, come to love granola or transformed her personality in any way. She has merely reopened deliberation. On the other hand, we want to allow that agents may change their preferences in response to changing evidence. But we saw that the relevant features of Good Granola and Mighty Muesli have not changed between t_{n-1} and t_n. On some ways of elaborating the method of considered default choice, the agent has acquired a flimsy piece of new evidence: that the store is under new management. But on other tellings, for example when redeliberation is triggered after a set number of shopping trips, no new evidence has been acquired.[14]

So far, we have seen that the considered default method opens the door to incoherence. Note next that the consistent strategy method may, in expectation, produce less incoherence in this and similar applications. Suppose that the agent consistently chooses using lexicographic choice, as in Section 3. Lexicographic choice cannot produce strict preference symmetries in the way illustrated above, because lexicographic choosers never change their choices without change of relevant evidence. And in many situations, lexicographic choice may be less vulnerable to preference-reversals of other kinds. Suppose that the agent regards quality as the most important attribute of cereals. Our lexicographic chooser will initially pick Mighty Muesli over Good Granola based on its superior quality. And as Mighty Muesli shrinks vastly in size and increases dramatically in price, our lexicographic chooser will continue to buy Mighty Muesli based on its superior quality.

[14]Of course, we may always restore coherence by insisting that *something*, such as the self-locating fact that it is now t_n triggered a preference change by providing direct evidence about the quality of the prospects X', Y'. But on many ways of filling out the example, this move is unmotivated and psychologically unconvincing. The agent may well tell us that she does not now (and did not at t_{n-1}) treat self-locating evidence as significantly relevant to her cereal choice. (Similar remarks apply to a move which differentiates the prospects X', Y' of receiving cereal at t_{n-1} from the prospects X'', Y'' of receiving cereal at t_n).

But the coherence of our lexicographic chooser comes at the cost of decision quality. She is coherent precisely because of her stubborn refusal to reverse her initial choice based on changes in decision-relevant variables such as cost and size. Whereas our considered default chooser eventually (and incoherently) changes her mind and purchases what is now a better cereal, the lexicographic chooser consistently continues to choose what is now a worse product, because there is no point at which she can spare the deliberative resources to recognize that it is worse. In this way, the example illustrates how a quality-coherence tradeoff can arise between the increased decision quality and decreased coherence gained from increasing the variety of decisionmaking strategies used.

Structurally similar examples can be used to illustrate an accuracy-coherence tradeoff, in which an increased variety of strategies for forming judgments produces, in expectation, more accurate but less coherent judgments. Keep the same example, but suppose now that you are shopping on behalf of your Aunt Edna. Each time, you ask yourself what your Aunt Edna would prefer. Here are two ways in which you might make this judgment.

First, you could employ a *think-rethink strategy* of remaking the judgment each time using a simple heuristic rule. For example, you could use the heuristic TTB, which we saw in Section 3 is structurally analogous to lexicographic choice. Second, you could employ a *think-retrieve-rethink* strategy. You would begin by using a more demanding heuristic or nonheuristic strategy on your first shopping trip to judge what Edna would prefer. On subsequent trips, you would search memory for a stored belief about Edna's preferences, and would accept this belief without redeliberation.[15] You would continue in this way until redeliberation was prompted, for example because enough time had elapsed or because evidence suggested redeliberation might be in order. Then you would deliberate again using a demanding strategy. The think-retrieve-rethink strategy is analogous to considered default choice.

As before, following the think-retrieve-rethink strategy opens the door to incoherence.

[15] This is similar to the initial retrieval step in the probabilistic mental models approach (Gigerenzer et al. 1991).

At t_{n-1} you would judge that Edna strictly prefers Mighty Muesli to Good Granola, and at t_n you would judge that Edna strictly prefers Good Granola to Mighty Muesli. This threatens to violate the requirement for judgments to be consistent:

> **(Judgmental Consistency)** For all collections $\{A_i\}$ of propositions, if $\wedge_i A_i$ is a contradiction, then you ought not (judge that A_1 and ... and judge that A_n).

On most tellings, the judgments that Edna would prefer Mighty Muesli at t_{n-1} but prefer Good Granola at t_n are inconsistent with the agent's background judgments about cereals and about Edna. For example, the agent may judge that Edna has not changed her cereal preference between t_{n-1} and t_n. And this judgment is inconsistent with her judgments about Edna's preferences at t_{n-1} and t_n.

Again as before, the think-rethink strategy avoids this type of incoherence because TTB cannot produce different judgments without a change in relevant information. As a result, the think-rethink strategy may outperform the think-retrieve-rethink strategy in expected coherence over this and similar judgment problems. But the think-rethink strategy gains in coherence precisely by refusing to change its mind, and continuing at t_n to make the false judgment that Edna would prefer what is now a much smaller and more expensive cereal. The think-rethink strategy purchases an increase in expected coherence by a decrease in expected accuracy. This illustrates how an accuracy-coherence tradeoff can arise from considerations of strategic variety.

In this section, we have seen that considerations of strategic variety can induce both quality-coherence and accuracy-coherence tradeoffs. For many agents, increasing the variety of cognitive strategies that they use will increase the expected accuracy of their judgments and quality of their decisions, but open the door to new forms of incoherence. In the next section, I argue that increasing cognitive complexity is a second way to induce quality-coherence and accuracy-coherence tradeoffs.

5 Coherence and cognitive complexity

It is often observed that nonhuman animals think and act more coherently than humans do (Searle 2001; Stanovich 2013). Whereas human judgment and decisionmaking is subject to many forms of probabilistic, logical and decision-theoretic incoherence (Kahneman et al. 1982; Gilovich et al. 2002), many of these errors are less-often found in animal cognition, and they are most frequently documented in relatively sophisticated animals (Krupenye et al. 2015; Marsh and Kacelnik 2002). This trend continues down to the limiting case of plant cognition, in which no credible incoherence has been documented (Schmid 2016).

What explains this striking negative correlation between complexity and coherence? A natural explanation is that there is often a *coherence-complexity tradeoff* in cognition (Stanovich 2013). By representing and processing information in more complex ways, we increase our risk of making incoherent judgments and decisions. This happens because more sophisticated processes can produce forms of incoherence which simpler rules avoid.

Why, then, would sophisticated creatures choose to employ more complex representations and processing rules when they could instead use simpler ones? A natural suggestion is that increasing cognitive complexity conduces to other cognitive goods, and in particular increases the expected accuracy of judgments and the expected quality of decisions.[16] If that is right, then the complexity-coherence tradeoff will often give rise to an accuracy-coherence tradeoff. In choosing to use complex cognitive processes, agents accept a heightened risk of incoherence in exchange for a better chance of making accurate judgments and good decisions.[17]

We can see how the complexity-coherence tradeoff arises by considering two factors

[16] As emphasized in Section 2, the claim is that complexity and accuracy often trade off, not that they always do. That is the lesson of less-is-more effects (Geman et al. 1992; Gigerenzer and Brighton 2009).

[17] One piece of corroborating evidence comes from accounts of the evolution of cognitive complexity, which emphasize that cognitive complexity arose in order to allow agents to cope with environmental complexity (Godfrey-Smith 2002; Powell et al. 2017). The ability of complex strategies to capture complex variation in problem contexts is, quite plausibly, a source of incoherence. Because complex processes can represent and respond to more complex arrays of problem features, there is more opportunity for us to react very differently to similar or identical situations by representing and processing these situations differently. On this point see Stanovich (2013).

which drive it. In fact, these factors are ubiquitous enough that we can illustrate both of them using examples already considered above.[18]

First, as creatures consider larger quantities or more complex types of information, they have an increased need to flexibly manage *information utilization* during decision-making. It becomes increasingly impossible to consider all relevant information during decisionmaking, so agents must decide which information to consider on a given occasion based on local task demands. This means that different bodies of information can be used to make judgments or decisions involving the same object on different occasions, creating an opportunity for incoherence.

By way of illustration, consider the turn from lexicographic to semilexicographic choice in Section 3. Semilexicographic choice differs from lexicographic choice precisely in its selective willingness to consider additional information. If two options are nearly tied along a given cue, semilexicographic choice moves on to consider additional cues rather than halting and making a decision. This is, in many contexts, an excellent way to decide when more information is needed, since the presence of near-ties suggests that the currently-favored option may not be overall better than its competitor.

However, it is precisely this feature of semilexicographic choice that gives rise to intransitive preferences. Additional information is consulted in determining the preference between Car A and Car B, as well as between Car B and Car C, because near-ties arise in both cases. But no additional information is consulted to determine the preference between Car A and Car C, allowing an intransitive cycle of preferences to be formed.

As we have seen, we could remove the risk of intransitivity by reducing our information utilization, retreating from semilexicographic choice back to lexicographic choice.[19] But we saw in Section 3 that this would plausibly come at the expense of expected deci-

[18]For brevity, I focus on the case of decisionmaking, since we have already seen how these examples can be reworked to reveal analogous phenomena in judgment formation.

[19]Of course, we could also block the issue by always considering all available information. However, that is often infeasible for bounded agents, just like it is often infeasible to draw all relevant inferences. More generally, it's not clear that there is any 'quick fix' that will reduce the risk of incoherence without substantial increases in effort or inaccuracy. Thanks to an anonymous referee for pushing me to address this possibility.

sion quality. The turn from lexicographic to semilexicographic choice reduces expected coherence by making flexible use of additional information, but this information is often decision-relevant, hence the reduction in expected coherence comes with an increase in expected decision quality.[20]

A second way that the complexity-coherence tradeoff arises is in *deliberative resource management*. As we employ more complex forms of processing, we need to increasingly ration deliberative resources by being selective about when complex processes are employed and how much effort is put into complex processes when they are used.[21] The problem is that rational management of deliberative resources involves starkly contrasting investments of deliberative resources into different tasks. When one and the same object features in two different tasks, we can produce different judgments or decisions involving that object, not because our evidence has changed but rather because of a change in how we deliberate.

Return again to the method of considered default choice from Section 4. Considered default choice involves using an effortful process to make the initial choice between Mighty Muesli and Good Granola, then continuing to make that same choice by default until contextual information suggests that redeliberation could be in order. As we saw, considered default choice can produce asymmetric patterns of strict preference because redeliberation may reverse our previous defaults without any change in the characteristics of Mighty Muesli and Good Granola.

We could eliminate this vulnerability by simplifying considered default choice in a striking way: never redeliberate. Using this strategy, if you buy Mighty Muesli once then you will continue buying it until one of the cereals is discontinued. This simplification

[20]A similar phenomenon is illustrated by humans' heightened vulnerability to attribute framing, in which our evaluations of an object are affected by how an attribute of that object is presented (Levin et al. 1998). On the prevailing account, attribute framing happens when rich, semantically-valanced mental representations trigger a valence-consistent shift in processes such as attention and memory retrieval (Jain et al. 2020; Levin et al. 1998). This need not always be irrational, insofar as semantic linkages between cognitive representations are valuable information which is often better used than ignored. But it does create the possibility for incoherence when oppositely-valenced representations are triggered by different presentations of the same decision problem. For other work on rational framing effects see (Bermúdez 2020).

[21]On operationalizing the notion of mental effort, see Shenhav et al. (2017).

removes the threat of incoherence in a blunt way, by refusing to invest more deliberative resources even when contextual information suggests that it is time to reevaluate your options. This is exactly the type of stubbornness that we took as evidence of reduced decision quality in a more complex method, the consistent strategy method in Section 4. By the same token, we should think that the stubbornness of default choice without redeliberation comes with a reduction in expected decision quality.[22]

In this section, we have seen two ways that a complexity-coherence tradeoff between cognitive complexity and coherence can be generated. First, using more and more complex information generates a need for selective information utilization policies, allowing different information to be used to evaluate the same object in different contexts. Second, employing complex processes creates a heightened need for deliberative resource management, allowing different judgments and decisions to be made about the same object when different types and quantities of deliberative resources are brought to bear on evaluating it. In both cases, we saw how the complexity-coherence tradeoff can give rise to an accuracy-coherence tradeoff, because appropriate use of more complex information and deliberative processes can increase expected accuracy and decision quality, even as it creates new opportunities for incoherence.

This completes my argument for the existence of an accuracy-coherence tradeoff in cognition. In the next section, I raise a puzzle about the possibility of an accuracy-coherence tradeoff. I use this puzzle to introduce an important distinction that will help us to locate the accuracy-coherence tradeoff within the structure of an agent's cognition.

---

[22]Strategies for deliberative resource management also contribute to the variety-coherence tradeoff across decision contexts. One of the major impetuses for applying different strategies to similar problems is that varying stakes or other changes in problem context make a differential investment of deliberative resources appropriate in each context. As we saw, shifts in strategy can be a source of incoherence, even if the shift is towards a more accurate and demanding strategy.

6 Locating the accuracy-coherence tradeoff

Sections 3-5 argued for the existence of accuracy-coherence and quality-coherence tradeoffs in cognition. Bounded agents are often forced to choose between structuring their cognition in ways most conductive to accuracy and decision quality or most conducive to coherence. In the special case of judgmental accuracy, this finding poses a puzzle. We began our discussion by noting that there are strong provable connections between the accuracy and coherence of belief states. These connections seem to place principled limits on the degree to which accuracy and coherence can trade off during cognition. This may suggest that the arguments for an accuracy-coherence tradeoff have been overstated. But where did we go wrong?

For concreteness, it will help to rehearse a few of the connections that I have in mind. Fix a scoring rule $\mathcal{I}$ measuring the inaccuracy of credal states and a divergence d measuring the distance between credal states. Say that credence function c *strictly* $\mathcal{I}$*-dominates* c' if in all possible worlds, c is less inaccurate than c'. Say that c *weakly* $\mathcal{I}$*-dominates* c' if in all worlds c is no more inaccurate than c', and in some world c is less inaccurate than c'. Accuracy dominance, at least in its strict form, is often taken as a strong sign of accuracy superiority.

One way to improve the coherence of a credence function is to move in a straight line towards a nearby coherent credence function. For any credence function c, a *d-nearest* coherent credence function c^* is any coherent credence function which lies at a minimal distance from c among the set of all coherent credence functions. Across a range of plausible candidates for $\mathcal{I}$ and d (Staffel 2020), it turns out that moving towards a nearest coherent credence function always produces a result strictly dominating the original function c:

> **(Linear Improvement)** If c is an incoherent credence function and c^* is a d-nearest coherent credence function to c, then for any $\lambda \in (0, 1)$ the credence function $\lambda c + (1 - \lambda)c^*$ strictly $\mathcal{I}$-dominates c.

In this sense, moving towards coherence is a good way to improve the accuracy of a credence function.

More generally, we can measure the *degree of incoherence* of a credence function c as the distance under d between c and a d-nearest coherent credence function. Say that credence function c is *more coherent than* c' just in case c has a lower degree of incoherence than does c'. Across a range of assumptions (Staffel 2020), we can establish a close relationship between dominance and degrees of coherence:

(Dominance-to-Coherence) For all credence functions c and c':

I If c weakly $\mathcal{I}$-dominates c', then c is not more incoherent than c'.

II If c strictly $\mathcal{I}$-dominates c', then c is less incoherent than c'.

Together, Dominance-to-Coherence and Linear Improvement suggest that the coherence and accuracy of credal states will not diverge very far. While these results stop short of showing that coherence and accuracy march in lock-step, it would be odd in view of these results if there turned out to be a strong negative correlation between the accuracy and coherence of credal states, considered in isolation.

How, then, could there be an accuracy-coherence tradeoff in cognition? The answer is that we went looking for this tradeoff in the wrong place. Herbert Simon held that the fundamental turn in the study of bounded rationality is the turn from *substantive* to *procedural* rationality (Simon 1976).[23] Theories of substantive rationality ask normative questions about the outcomes of deliberation such as credences and preferences, without considering the processes of deliberation that produced them. Theories of procedural rationality pay explicit attention to the processes of deliberation through which these outcomes are produced.

[23]This terminology is unfortunate, because Simon's distinction between substantive and procedural rationality is independent of the coherence-based theorist's distinction between substantive and structural rationality. In this section, I use 'substantive rationality' in Simon's sense. In Section 7, I use it in the coherentist's sense.

Most theorists in the bounded tradition have followed Simon in holding that substantive rationality is a misleading lens through which to study bounded rationality. The problem for theories of substantive rationality is that if we only study the outcomes of deliberation, we will neglect important cognitive bounds such as the costs of cognition and limitations on our ability to make certain inferences and decisions. If we do not model the process of deliberation, these bounds will fall outside of our models.

As a fact about substantive rationality, it may be quite right to say that there is a strong correlation between the coherence and accuracy of credal states, considered in isolation from the processes that produced them. But when we zoom out to examine the processes of deliberation through which credences are produced, the correlation between accuracy and coherence breaks down. In order to increase the coherence of our credences, we need to make specifiable changes to the cognitive processes through which these credences are produced. The lesson of Sections 3-5 is that after a point, these changes begin to have unsavory consequences for the accuracy of an agent's credences. These consequences are revealed by examining the ways in which an agent's cognitive processes would have to change in order to produce more coherent judgments, together with the accuracy-costs of employing these processes.

The resulting tension between accuracy and coherence is not revealed by results such as Linear Improvement and Dominance-to-Coherence which deal directly with the accuracy and coherence of credence functions but make no mention of the processes of deliberation that produced them. The accuracy-coherence tradeoff appears only at the level of procedural rationality. The fact that important tradeoffs such as the accuracy-coherence tradeoff are only revealed at the procedural level illustrates the importance of studying procedural rationality.

In this section, we began with a puzzle. When credences are considered in abstraction from the processes that produced them, we recover tight connections between coherence and accuracy. This raised the puzzle of how to make room for an accuracy-coherence tradeoff in cognition. The answer is that the accuracy-coherence tradeoff is felt most strongly at the procedural level of deliberation procedures rather than the substantive level of beliefs and preferences that result from these procedures. For exactly this reason, bounded rationality theorists have held that theorizing about bounded rationality occurs mostly at the procedural level. What implications does this have for the normative place of coherence in bounded rationality?

7 Normative implications

In the unbounded case, disputes about the normative status of coherence requirements are not primarily extensional disputes. Coherence-based normative theories hold that rational agents must have coherent beliefs and preferences, because coherence requirements are fundamental requirements of rationality. Their opponents agree that rational agents must have coherent, or nearly-coherent attitudes, but take this fact to be derivative on other rational requirements, such as the requirement to hold evidentially supported beliefs (Kolodny 2005). What is at issue is not, for the most part, what rationality requires but whether the requirements in question are fundamental or derivative. Even distinguishing between structural and substantive rationality or between oughts and rational requirements will not bring out significant extensional disagreement, for it is widely agreed that unbounded agents ought, and are substantively rationally required to hold mostly coherent attitudes.[24]

It looks tempting to extend coherence requirements to bounded agents by holding that bounded agents are rationally required to be as coherent as possible, given their limitations. But now we arrive at an extensional dispute, because for bounded agents there is an accuracy-coherence tradeoff between the accuracy and coherence of their judgments as well as a quality-coherence tradeoff between the quality and coherence of their preferences or decisions. Now it is tempting to hold that what agents ought to do,

[24]Some authors go further and target coherence norms in their application to belief, or in the life of unbounded agents (Arkes et al. 2016; Field forthcoming). An advantage of the present approach is that it does not require us to go this far.

and what substantive rationality requires them to do, is not to pursue coherence at all costs, but rather to adopt cognitive strategies which sacrifice some amount of expected coherence for the sake of improvements in accuracy and decision quality. This leaves coherence-based theories with three options.[25]

First, they can say that the requirement for bounded agents to be as coherent as possible is a requirement of structural rather than substantive rationality, or a rational requirement rather than a statement about what ought to be done. The ensuing dialectic is familiar, but the tone has changed. It is now conceded that what agents ought to do or what substantive rationality requires them to do is not always, or even perhaps for the most part to be as coherent as possible. This means that there is a central normative sense in which many agents may be uncriticizable, or even praiseworthy for failures to pursue coherence.

Second, coherence-based theories may not lean on these distinctions and may instead make the stronger claim that what agents ought to do and what they are rationally required to do is to pursue coherence at all costs. But now Kolodny's (2005) challenge returns with a vengeance: why be rational? In the present context, the challenge strengthens to the following: why be rational, instead of structuring our cognitive lives in ways expected to produce more accurate judgments and better decisions? This is a tough question to answer, and it threatens to tell against coherence-based accounts.

Finally, coherence-based theories could concede that overall rational requirements are partially informed by considerations of accuracy and decision quality, which trade off against coherence.[26] This would involve denying that bounded agents are rationally required to be as coherent as possible. This move raises interesting normative questions, of which I highlight two examples below.[27]

---

[25]How about a fourth option: once you've settled on a broad class of cognitive strategies, you should strive to implement those strategies as coherently as possible? In general, I'm not opposed to emphasizing coherence in later steps of strategy-selection, although I suspect it might be a bit strong to give coherence complete priority over other goals at later stages of strategy selection. Thanks to an anonymous referee for pushing me on this point.

[26]Or relatedly, they could follow Broome (2013) in letting rationality be exhausted by coherence, but treating rationality as one of many considerations bearing on how we ought to cognize.

[27]This may also put pressure against attempts to account for heuristic rationality on grounds of accuracy,

First, classic experiments from the heuristics and biases program show that heuristic strategies can sometimes produce incoherent attitudes. Defenders of heuristic cognition counter that these occasional biases are the predictable result of rational heuristic procedures for judgment and decisionmaking. Although heuristics occasionally produce incoherent judgments, they may nonetheless strike the best balance between accuracy and coherence in cognition. If this is right, then in what sense, if any, are agents rationally criticizable for forming incoherent attitudes as a result of heuristic cognition? Might we go so far as to say that there is no sense in which they are rationally criticizable, for they would have been irrational to cognize in any other way?[28]

Second, if accuracy and decision quality are to trade off against coherence in determining what rationality requires, then how are these considerations to be weighed against one another?[29] Presumably, coherence-based theories want to avoid a view on which accuracy and decision quality should always outweigh considerations of incoherence, but it is not easy to see how this tradeoff is to be made. Moreover, ardent defenders of an accuracy-centered approach to epistemology may claim that accuracy should always take precedence over coherence in judgment, and many, such as utilitarians, might claim that decision quality should always take precedence over coherence in decisionmaking. This raises the challenge of saying how and why coherence should be allowed to trade off against accuracy and decision quality.

Summing up, we have seen that bounded agents face accuracy-coherence and quality-coherence tradeoffs, in which the accuracy and quality of cognitive processes trades off against their expected coherence in many cognitive contexts. These tradeoffs arise because subsidiary factors such as strategic variety and cognitive complexity often increase expected accuracy and quality, but decrease expected coherence. Taking these tradeoffs

---

without inducing downstream revisions to familiar epistemological principles (Karlan forthcoming).

[28]For recent philosophical attempts to vindicate cognitive biases, see Hedden (2019), Icard (2018), Morton (2017), and Polonioli (2013).

[29]See Worsnip (forthcoming) for discussion of the difficulty of weighing coherence against substantive considerations such as accuracy in deliberation.

seriously highlights the need for theories of bounded rationality to attend to procedural rationality, since it is at this level that the tradeoff appears. These tradeoffs also restrict the form of defensible coherence-based approaches to rationality, and raise interesting normative questions that are deserving of further study.

References

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

Armstrong, J. Scott and Graefe, Andreas. 2012. "Predicting elections from the most important issue: A test of the take-the-best heuristic." *Journal of Behavioral Decision Making* 25:41–48.

Bermúdez, José. 2020. *Frame it again*. Cambridge University Press.

Broome, John. 2013. *Rationality through reasoning*. Wiley.

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.

De Bona, Glauber and Staffel, Julia. 2017. "Graded incoherence for accuracy-firsters." *Philosophy of Science* 84:189–213.

—. 2018. "Why be (approximately) coherent?" *Analysis* 78:405–415.

Field, Claire. forthcoming. "Embracing incoherence." In Nick Hughes (ed.), *Epistemic dilemmas*, forthcoming. Oxford University Press.

Fishburn, Peter. 1974. "Lexicographic orders, utilities and decision rules: A survey." *Management Science* 20:1442–1471.

Geman, Stuart, Bienenstock, Elie, and Doursat, René. 1992. "Neural networks and the bias/variance dilemma." *Neural Computation* 4:1–58.

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

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

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

Gigerenzer, Gerd, Hoffrage, Ulrich, and Kleinbölting, Heinz. 1991. "Probabilistic mental models: A Brunswikian theory of confidence." *Psychological Review* 98:506–528.

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

Gilovich, Thomas, Griffin, Dale, and Kahneman, Daniel (eds.). 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Godfrey-Smith, Peter. 2002. "Environmental complexity and the evolution of cognition." In Robert Sternberg and James Kaufman (eds.), *The evolution of intelligence*, 223–49. Lawrence Erlbaum.

Hedden, Brian. 2019. "Hindsight bias is not a bias." *Analysis* 79:43–52.

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

Jain, Gaurav, Gaeth, Gary, Nayakankuppam, Dhananjay, and Levin, Irwin. 2020. "Revisiting attribute framing: The impact of number roundedness on framing." *Organizational Behavior and Human Decision Processes* 161:109–19.

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

Johnson, Eric J. and Goldstein, Daniel. 2003. "Do defaults save lives?" *Science* 302:1338–1339.

Kahneman, Daniel, Slovic, Paul, and Tversky, Amos (eds.). 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Karlan, Brett. forthcoming. "Reasoning with heuristics." *Ratio* forthcoming.

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

Krupenye, Christopher, Rosati, Alexandra G., and Hare, Brian. 2015. "Bonobos and chimpanzees exhibit human-like framing effects." *Biology Letters* 11:20140527.

Levin, Irwin, Schneider, Sandra, and Gaeth, Gary. 1998. "All frames are not created equal: A typology and critical analysis of framing effects." *Organizational Behavior and Human Decision Processes* 76:149–188.

Manzini, Paola and Mariotti, Marco. 2012. "Choice by lexicographic semiorders." *Theoretical Economics* 7:1–23.

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

Marsh, Banabay and Kacelnik, Alex. 2002. "Framing effects and risky decisions in starlings." *Proceedings of the National Academy of Sciences* 99:3352–3355.

Martignon, Laura and Hoffrage, Ulrich. 2002. "Fast, frugal and fit: Simple heuristics for paired comparison." *Theory and Decision* 52:29–71.

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

Payne, John, Bettman, James, and Johnson, Eric. 1988. "Adaptive strategy selection in decisionmaking." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:534–552.

Polonioli, Andreas. 2013. "Blame it on the norm: The challenge from 'Adaptive rationality'." *Philosophy of the Social Sciences* 44:131–150.

Powell, Russell, Mikhalevich, Irina, Logan, Corina, and Clayton, Nicola S. 2017. "Convergent minds: The evolution of cognitive complexity in nature." *Interface Focus* 7:20170029.

Rysiew, Patrick. 2008. "Rationality disputes - psychology and epistemology." *Philosophy Compass* 3:1153–1176.

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* 8:347–355.

—. 2002. "Measuring Incoherence." *Sankhya: The Indian Journal of Statistics* 64:561–587.

—. 2003. "Measures of incoherence: How not to gamble if you must." In José Bernardo (ed.), *Bayesian statistics*, 385–401. Oxford University Press.

Schmid, Bernhard. 2016. "Decision-making: Are plants more rational than animals?" *Current Biology* 26:R675–R678.

Searle, John. 2001. *Rationality in action*. MIT Press.

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.

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

Staffel, Julia. 2015. "Measuring the overall incoherence of credence functions." *Synthese* 192:1467–1493.

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

Stanovich, Keith. 2013. "Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice." *Thinking and Reasoning* 19:1–26.

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

Sturm, Thomas. 2012. "The 'rationality wars' in psychology: Where they are and where they could go." *Inquiry* 55:66–81.

Tversky, Amos. 1969. "Intransitivity of preferences." *Psychological Review* 76:31–48.

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

Worsnip, Alex. forthcoming. "Making space for the normativity of coherence." *Noûs* forthcoming.

Zynda, Lyle. 1996. "Coherence as an ideal of rationality." *Synthese* 109:175–216.