The complexity-coherence tradeoff in cognition

David Thorstad

Forthcoming in *Mind*, please cite published version

Abstract

I present evidence for a systematic complexity-coherence tradeoff in cognition. I show how feasible strategies for increasing cognitive complexity along three dimensions come at the expense of a heightened vulnerability to incoherence. I discuss two normative implications of the complexity-coherence tradeoff: a novel challenge to coherence-based theories of bounded rationality and a new strategy for vindicating the rationality of seemingly irrational cognitions. I also discuss how the complexity-coherence tradeoff sharpens recent descriptive challenges to dual process theories of cognition.

1 Introduction

Here is a puzzling fact.[1] It is widely agreed that humans are the least coherent creatures on Earth. There are well-documented circumstances in which humans violate nearly every requirement of coherent belief, credence, preference or choice ever proposed (Kahneman et al. 1982; Shafir and LeBoeuf 2002). In nonhumans, incoherence is more rarely observed, and then often in the most complex creatures such as primates (Krupenye et al. 2015) and starlings (Schuck-Paim 2002). An incoherent rat is a noteworthy scientific finding (Sweis et al. 2018). And in the least complex creatures, incoherence is rarely found.[2] In the limiting case of plant cognition, no incoherence has ever been observed (Schmid 2016).[3] Why would the most complex creatures on Earth also be the least coherent? This finding is especially puzzling on coherence-based theories of rationality (Staffel 2020; Zynda 1996). On these views, we must explain why the most complex creatures on Earth would also be the least rational, and why they would not choose to be more rational by choosing to cognize in simpler ways.

The inverse relationship between complexity and coherence is often noted, but rarely explained. For example, Alison Gopnik wonders: 'Why are grown-ups often so stupid about probabilities when even babies and chimps can be so smart?' (Gopnik 2014). And

[1] I am grateful to Zachary Gabor, Jeremy Goodman, Uriah Kriegel, Toby Newberry, Rhys Southan, Teru Thomas, and audiences at the Australasian Association of Philosopy, Berkeley Bounded Rationality Conference, Formal Epistemology Workshop, and the Formal Methods Group at King's College, London for comments and discussion.

[2] Perhaps Shafir (1994) and Dawkins and Brockmann (1980) are credible examples of incoherence in honeybees and wasps, although in such cases both the nature of coherence and the interpretation of experimental results become controversial (Arkes and Ayton 1999).

[3] Some readers may not be sympathetic to the idea of plant cognition. These readers are invited to picture the simplest creature to which they are willing to attribute cognitive states. How many times has such a creature been observed to think or act incoherently?

John Searle (2001) begins his criticism of received economic models of rationality by noting that chimpanzees often perform at least as well as humans on classical models. But for their part, neither Gopnik nor Searle explains why it is that complex creatures, despite their cognitive advantages, should be less coherent than simpler creatures.

One possibility is that the inverse relationship between coherence and cognitive complexity is a coincidence. But if it is a coincidence, it is a strikingly consistent one. My point of departure is a recent suggestion that the inverse relationship between complexity and coherence is not a coincidence, but rather part of a systematic complexity-coherence tradeoff in cognition (Stanovich 2013; Thorstad forthcoming).[4] The most natural way to explain why complex creatures tend to be less coherent than simpler creatures is to say that complex cognitive processes tend to produce more incoherent results than simpler processes do. This paper argues that the natural explanation is correct.

More precisely, my aim in this paper is to do three things. First, I clarify what it means to speak of a complexity-coherence tradeoff in cognition (§2). Second, I argue that the complexity-coherence tradeoff often obtains and catalog three of the factors driving the complexity-coherence tradeoff: procedural complexity (§3), aspiration adaptation (§4) and informational complexity (§5). Finally, I draw out normative and descriptive implications of the complexity-coherence tradeoff and sketch directions for future work (§6).

In particular, I discuss how agents should choose between the competing goals of complexity and coherence in cognition (§6.1). Then I show how the complexity-coherence tradeoff generates a novel challenge to coherence-based theories of bounded rationality (§6.2). Once we see that the pursuit of coherence often comes at the expense of complex cognition, it becomes relatively less attractive to privilege coherence over competing cognitive goals, and more attractive to offer error theories for coherence-based approaches to bounded rationality.

I also show how the complexity-coherence tradeoff opens new avenues for vindicatory epistemology (Thorstad forthcoming b), the project of vindicating rationality of seemingly irrational cognitions (§6.3). In particular, the complexity-coherence tradeoff opens new avenues for vindicating intransitive preferences, ordering effects, and framing effects, as well as for defending the rationality of heuristic cognition and strategies which are sensitive to the format in which information is presented.

Finally, I show how the complexity-coherence tradeoff sharpens a recent line of descriptive attack against dual process theories of cognition, which questions the explanatoriness or even the well-definedness of dual process theories by problematizing the central dichotomies used to introduce and apply dual process theories (§6.4). Typical versions of dual process theory hold that complex Type 2 processes tend to produce more coherent results, whereas simpler Type 1 processes tend to produce less coherent results, but the complexity-coherence tradeoff suggests precisely the opposite pattern, putting pressure on a central contention and explanatory application of dual process theories. I show how this discussion challenges recent descriptive applications of dual process theorizing. I also explore normative implications of this discussion, including the failure of some recent debunking explanations for cognitive biases and nonconsequentialist moral intuitions, as well as the importance of recent challenges to nudging.

[4] Morton (2010) also anticipates this suggestion in some respects.

2 Clarifying the target

What does it mean to say that there is a complexity-coherence tradeoff in cognition? Five remarks will help to clarify this claim.

First, we need to distinguish tradeoffs between features of attitudes from tradeoffs between features of the cognitive processes that produce them (Parfit 1984; Railton 1984; Thorstad forthcoming b). Most classic tradeoffs in the theory of bounded rationality are understood as tradeoffs between features of cognitive processes, and the complexity-coherence tradeoff is no exception.[5] To posit a complexity-coherence tradeoff in cognition is to say that agents must choose among a feasible range of cognitive processes, and that the most complex of these processes are not always, in expectation, the processes which produce the most coherent attitudes.[6] The complexity-coherence tradeoff between features of cognitive processes is not to be equated with any claim about attitudes, such as the claim that coherence and complexity are anti-correlated features of attitude sets, specified in isolation from the processes that produced them. That is a distinct claim which would require separate treatment.

Second, most classic tradeoffs in the bounded tradition hold often, but not always.[7] For this reason, the most important research project is to identify the factors which drive the presence or absence of any given tradeoff (Thorstad forthcoming; Todd and Gigerenzer 2012). To posit a complexity-coherence tradeoff in cognition is to say that in many situations, the most complex feasible cognitive processes are not always, in expectation, those that produce the most coherent attitudes.[8] My project in this paper is to identify some of the many factors which may drive the complexity-coherence tradeoff. I focus on three factors: procedural complexity, aspiration adaptation, and informational complexity.

Third and relatedly, in talking of a complexity-coherence tradeoff we must restrict attention to a range of feasible strategies that may be reasonably implemented by agents with limited capacities. The claim is that among these feasible strategies, the most complex cognitive strategies come apart from the most coherent strategies. In many situations, I do not want to deny that there exists some much more complex strategy that would, if implemented, lead only to coherent attitudes. For example, in finite choice settings agents could simply list all pairwise choices and form preferences consistent with their previously formed pairwise preferences. My claim is rather that within a feasible range

[5] These include the accuracy-effort tradeoff (Johnson and Payne 1985), speed-accuracy tradeoff (Heitz 2014) and bias-variance tradeoff (Geman et al. 1992; Gigerenzer and Brighton 2009).

[6] We might perhaps extend the complexity-coherence tradeoff to other features of cognition that are not cognitive processes, and which are not selected by agents, such as the cognitive architectures adapted through biological evolution. Indeed, Okasha (2018) and Spurrett (2021) argue that the evolutionary factors favoring complex cognition do not always favor coherence. But my interest in this paper is only with the cognitive processes that agents select during their lifetimes.

[7] For example, the accuracy-effort tradeoff reverses when the bias-variance dilemma begins to bite (Gigerenzer and Brighton 2009; Wheeler 2020).

[8] How often does the complexity-coherence tradeoff obtain? The short answer is: often, but not always. The long answer is: when the motivating stories in Sections 3-5, or other stories like them, hold and are not counteracted by significant countervailing factors. The longest answer is that as with existing tradeoffs such as the accuracy-effort and bias-variance tradeoffs, mapping the shape of the complexity-coherence tradeoff is a research program, not something to be settled in an individual paper. A good start is to identify factors driving the complexity-coherence tradeoff and to say roughly when and why we might expect them to drive the tradeoff. That is what I aim to do in this paper.

of complexity, increasing complexity often comes at the expense of coherence.

Fourth, the notion of coherence raises interpretive difficulties. Not all traditions agree on the requirements of coherence, and some differ also on whether coherence is the right umbrella term to pick out the requirements of interest. For the most part, I concentrate on simple requirements of coherence, such as intransitivity, symmetry and reflexivity of strict preference, which are shared across most competing views. However, it is also important to show how the complexity-coherence tradeoff applies to broader notions of incoherence, such as classic behavioral biases. For this reason, I include a case study of framing effects.

Fifth, the notion of complexity is equally fraught. One problem is that complexity is studied through a number of different approaches, including complex systems theory (Ladyman and Wiesner 2020), information theory (Shannon 1948), psychology (Liu and Li 2012) and behavioral economics (Oprea 2020). These approaches are not always directly comparable, and when they are comparable they do not always agree. A second problem is that there is substantial disagreement within approaches (Ladyman and Wiesner 2020). For example, a range of conflicting information-theoretic complexity criteria have been defended, including Shannon entropy (Shannon 1948), Kolmogorov complexity (Kolmogorov 1965), logical depth (Bennett 1988), effective complexity (Gell-Mann 1995), and statistical complexity (Crutchfield and Young 1989). I address this problem by focusing on a wide variety of complexity notions, doing my best to characterize these notions in a theory-light way that can translate into several different disciplinary approaches. These notions include procedural complexity (§3), state complexity (§4), and informational complexity (§5). At the same time, I recognize that the exact extension of the complexity-coherence tradeoff will be sensitive to views about complexity, just as it is sensitive to views about coherence. It would be an interesting project for future work to map the contours of the complexity-coherence tradeoff against varying notions of complexity and coherence.

Summing up, the complexity-coherence tradeoff is in the first instance a claim about cognitive processes. My claim is that the complexity-coherence tradeoff occurs often, not always, and in particular that this tradeoff emerges once we restrict attention to a range of feasible strategies. I map the complexity-coherence tradeoff across a range of complexity and coherence concepts, doing my best to provide a selection of examples that will satisfy most theorists. With these clarifications in mind, let us begin with an example designed to illustrate how a complexity-coherence tradeoff could arise.

3 Lexicographic choice and procedural complexity

Begin with *procedural complexity*, the quantity and complexity of processing steps involved in executing a cognitive process.[9] Feasible ways of increasing procedural complexity often introduce new opportunities for incoherence. Because there are more moving parts, there are more opportunities for these parts to move in opposite directions. This is, plausibly, one reason why more complex creatures tend to be less coherent: they can and do execute more complex cognitive processes, with more opportunities for incoherence to

[9] Procedural complexity is recognized as a type of complexity in many leading taxonomies. For example, Bonner (1994) classifies processing complexity as one of three types of task complexity, and Liu and Li (2012) specify seven *complexity contributory factors* of processes which increase complexity.

result. When this happens, procedural complexity will generate a complexity-coherence tradeoff: feasible increases in procedural complexity decrease the expected coherence of the resulting attitudes.

To illustrate, suppose you are deciding between several vacation destinations. One way you might make this choice is by *tallying*. Tallying instructs you to retrieve some number n of features that you care about, such as warm weather and the availability of tennis courts.[10] For simplicity, we will suppose that you retrieve $n = 2$ features, though it would be more sensible to retrieve 5 or 10 features. Let's take the simplest case in which all features are binary: for example, either a destination has warm weather (feature value 1) or it does not (feature value 0). For each feature, you then compute the *tally* of positive features among the n features retrieved. With binary features, this amounts to summing feature values. You would then halt choice and settle on the option with highest tally.[11]

Tallying is a sensible way to make many decisions. Indeed, under many conditions tallying meets or exceeds the performance of much more demanding processes, such as linear regression (Dawes and Corrigan 1974; Dawes 1979). However, tallying has a drawback: it settles near ties in favor of the option with highest tally. When cognitive resources are not especially tight, it may be a feasible improvement to consider more features of nearly-tied options in order to make a better-informed choice.

Let *near-tallying* be a process which coincides with tallying, except in cases where two or more options have final tallies within m of the best tally. For simplicity, we will take $m = 1$. In this case, near-tallying instructs agents to retrieve another n features, then compute the tally of each of the nearly-tied options. If one option now leads the tally by more than m, that option is chosen. Otherwise, another n features are examined and choice repeats as before.

However, near-tallying is strictly less coherent than tallying.[12] Many theorists accept as a minimal requirement of coherence that strict preferences should be transitive:

> **(Transitivity of Strict Preference)** For all agents S and options o, o', o'' if $o \succ_S o'$ and $o' \succ_S o''$ then $o \succ_S o''$.

Tallying always satisfies transitivity, but near-tallying may not. Suppose our near-tallying vacationer is confronted with the three vacation options in Table 1. Between Option A and Option B, our near-tallier chooses feature A after examining all 8 features. Between Option B and Option C, our near-tallier chooses option B after examining 6 features. Between Option A and option C, our near-tallier chooses option C after examining 2 features. This looks to reveal a collection of intransitive strict preferences: Option A is strictly preferred to Option B, Option B is strictly preferred to Option C, and Option C is strictly preferred to Option A.[13]

[10] We also need to specify the order in which features are retrieved. Features might be ordered by importance, randomly, or in some other order.

[11] In the case of ties, you would become indifferent between each item with maximal tally. Choice would be resolved through your favorite procedure for indifferent choice.

[12] The textual discussion shows that near-tallying, but not tallying, violates the transitivity of strict preference. In the other direction, tallying could not introduce any incoherence not already present in near-tallying, since tallying is a type of near-tallying.

[13] Some authors might hold that choice reveals only weak preference in this case. While this is not my view, those who hold it are welcome to enrich the option space with mild sweetenings of each option in order to show the intransitivity to be strict.

	Option A	**Option B**	**Option C**
Feature 1	0	1	1
Feature 2	0	0	1
Feature 3	1	1	0
Feature 4	1	1	0
Feature 5	1	1	0
Feature 6	1	0	0
Feature 7	1	0	0
Feature 8	1	0	0

Table 1: Near-tallying applied to three vacation options

Suppose now that our vacationer can feasibly implement two strategies: tallying or near-tallying. She is then faced with a complexity-coherence tradeoff. Opting for near-tallying yields a feasible increase in complexity, since near-tallying adds potential additional processing steps to tallying. However, opting for near-tallying decreases the expected coherence of the attitudes that will result. If this is right, then procedural complexity can be seen as a first factor driving the complexity-coherence tradeoff, with feasible and potentially desirable increases in procedural complexity leading to a heightened risk of incoherence.

4 K-phase satisficing and aspiration adaptation

A general difficulty in theorizing about complexity is that many examples rely on unformalized notions of complexity. This restricts the range of examples we can consider to those where one process is in a clear and intuitive sense more complex than another. For example, we held that semilexicographic choice has higher procedural complexity than lexicographic choice because semilexicographic choice is a strict extension of lexicographic choice that adds a novel tie-breaking step.

To expand our diet of examples, it will help to work with a formalized notion of complexity. This requires fixing a specific cognitive architecture in which processes can be implemented and settling on a formal measure of complexity. Both the choice of cognitive architecture and formal complexity measure will be controversial, so it is best to supplement such discussions with multiple models, as well as with alternative routes to a complexity-coherence tradeoff. In this section, I begin that effort by representing cognitive architecture using finite automaton theory, an approach popular in economics and computer science (Oprea 2020; Rubinstein 1986; Salant 2011).

Intuitively, increasing the number of states that an automaton can occupy increases the complexity of the automaton, but creates new opportunities for incoherence. Whereas a simple 1-state automaton always reacts to stimuli in the same way, more complex automata may treat stimuli in different ways depending on their state. This section illustrates a setting in which there is a clear tradeoff between the number of automaton states and the coherence of the automaton's decisions, then re-interprets that tradeoff in terms of aspiration adaptation, a special kind of learning.

An *automaton* A takes as input ordered lists L from a domain D and chooses an element

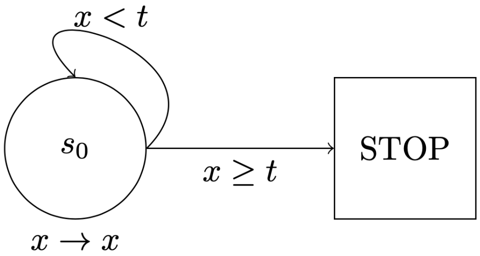

Figure 1: Satisficing

of the list L. An automaton $A = (S, s_0, g, f)$ has four components. S is a set of potential states the automaton can occupy, with s_0 the automaton's initial state. The automaton moves through list L one item at a time. The *transition function* $g : S \times D \longrightarrow S \cup \{\textsc{stop}\}$ tells A whether to transition into a new state or halt, upon observing list element $x \in D$ while in state $s \in S$. When halting, the *output function* $f : S \times D \longrightarrow D$ tells A which element of D to choose according to its previous state and last-observed input.

Cognitive processes can be studied by considering the *choice functions* $c : L(D) \longrightarrow D$ they implement, taking as input ordered lists of alternatives from D and returning a chosen element of the list. But automata also implement choice functions. An automaton A *implements* a choice function c just in case A and c return the same output on all lists in $L(D)$. This allows us to study the complexity of choice functions by studying the complexity of the automata that implement them. Many complexity notions are possible here (Oprea 2020), but one of the most studied is state complexity (Rubinstein 1986; Salant 2011). The *state complexity* of a choice rule is the minimal number of states required to implement it in a finite automaton.

For example, consider *satisficing* (Figure 1). In this context, satisficers fix a utility threshold t and choose the first element of L with utility t or higher. Satisficing has state complexity one, as it can be implemented by a single-state automaton. The transition function, represented by arrows between states, tells the automaton to stop once an option with utility t or higher is observed, and the output function, drawn beneath states, says to choose that option.[14]

By contrast, utility maximization has state complexity $|D| - 1$, one less than the cardinality of the option space (Figure 2). One way to implement utility maximization is to order the elements $x_1, \ldots, x_N$ of D by increasing utility, using states $s_1, \ldots, s_{N-1}$ to 'record' when a non-maximal element of D has been seen. The transition function $s(i, x_j) = max(i, j)$ shifts to a higher state once a better element is seen, except that $s(i, x_N) = \textsc{stop}$, halting if the best-possible element has been found. The output function chooses the best observed element once all list elements have been exhausted.[15] Regrettably, we can prove that no automaton with fewer than $|D| - 1$ states implements utility maximization (Salant 2011).

In many circumstances, utility maximization may be infeasibly complex. To borrow an example from Peter Bossaerts and Carsten Murawski (2017), the process of choosing a utility-maximizing basket of items out of a small grocery store stocking 1,000 items has state complexity on the order of 10^{301}, more than 10^{220} times the estimated number of atoms in the universe. In such cases, agents may seek a compromise between satisficing and

[14] That is, $g(s_0, x) = \textsc{stop}$ if $u(x) \geq t$ and $g(s_0, x) = s_0$ otherwise, with $f(s_0, x) = x$.

[15] I.e. $s(i, x)$ returns x if $u(x) > u(x_i)$ and otherwise returns x_i.

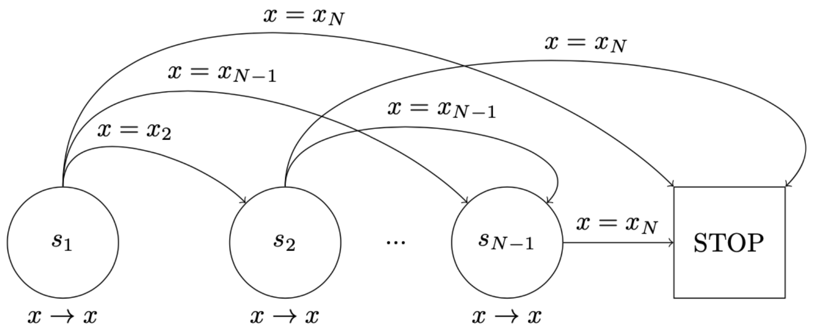

Figure 2: Utility maximization

utility maximization by designing choice processes with state complexity strictly between 1 and $|D| - 1$.

However, increasing state complexity within this range can lead to incoherence. Say that $L' < L$ if L' is a *sublist* of L in the sense that L' results from L by removing some elements of L. One common coherence requirement is the Independence of Irrelevant Alternatives: whatever is worth choosing from a list is still worth choosing from a sublist of the original list.

> **Independence of Irrelevant Alternatives (IIA)** If $x \in L' < L$ and $x = c(L)$ then $x = c(L')$.

Many authors hold that it would be incoherent to violate *IIA* by preferring x from the list L but not from a smaller list L'. After all, x has not changed, and the agent has not been offered any new alternatives to x, so it is hard to see how the agent could coherently decide to reject x from the smaller list.

Both satisficing, a 1-state process, and utility maximizing, a $|D|-1$-state process, satisfy IIA. But with state complexity strictly between 1 and $|D|-1$, the story is different. Suppose you face the *choice design problem* of adopting a choice rule, subject to the constraint that its state complexity be no more than K. And suppose we make a small structural assumption about how lists are generated: items are drawn one at a time from D by a probabilistic process P which has nonzero probability of picking each item from D. The process then halts with some constant probability c and otherwise generates another list item. Under these assumptions, we can prove that for $1 < K < |D| - 1$, the expected utility-maximizing solution to the choice design problem is *K-phase satisficing* (Salant 2011).

Informally speaking, K-phase satisficing is an agglomeration of K satisficing agents with increasingly demanding satisficing thresholds (Figure 3). Each threshold t_i beyond the first corresponds to a 'pivotal alternative' a_i that raises the threshold to t_i unless the threshold is already higher. This allows the agent to learn from experience that a more demanding threshold is appropriate.

Formally, K-phase satisficing begins with an initial threshold t_0 and a sequence of $K-1$ *pivotal alternatives* $a_1, \ldots, a_{K-1}$ from D. The pivotal alternatives are chosen so that $u(a_i) = t_i$, generating a sequence of increasing thresholds $t_0 < t_1 < \cdots < t_{K-1}$. The agent has states $s_0, \ldots, s_{K-1}$ corresponding to the satisficing thresholds t_i.

Choice proceeds as follows. When observing a non-terminal list element x in state i, if x is a pivotal alternative a_j then the agent shifts to state j if $j > i$, adjusting her satisficing

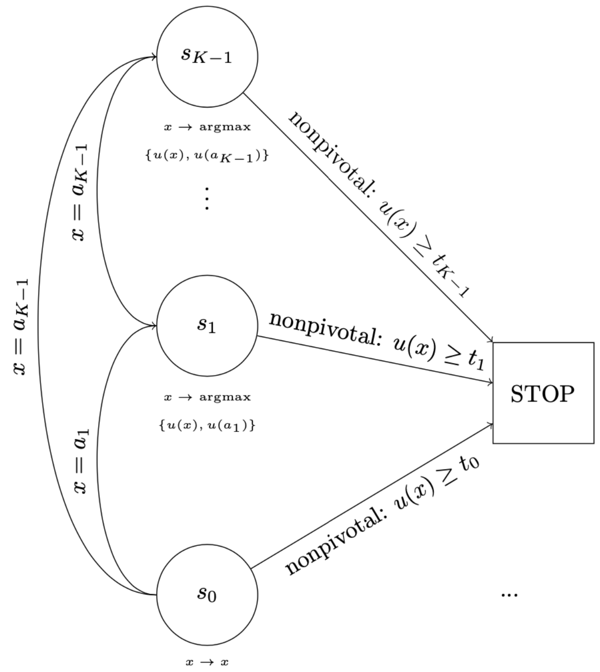

Figure 3: K-phase satisficing

threshold upwards to t_j. If x is non-pivotal, the agent satisfices with threshold t_i, halting with the choice of x if $u(x) \geq t_i$ and otherwise examining the next list element. In the special case that x is a terminal list element, the agent makes a forced choice between x and her currently favored alternative a_i, choosing x just in case $u(x) > u(a_i)$.

Despite its optimality, K-phase satisficing has a problem. For $1 < K < |D| - 1$, K-phase satisficing violates IIA. To see this, let L be the list $x_1x_2x_3$ and L' be the sublist x_2x_3. Let x_1 but not x_2 be a pivotal alternative, with $t_0 < u(x_2) < u(x_1) < u(x_3)$. Then K-phase satisficing selects x_3 from L, since x_1 raises the choice threshold above $u(x_2)$. But K-phase satisficing selects x_2 rather than x_3 from the sublist L', since x_1 is no longer around to raise the choice threshold above t_0. This is a violation of IIA.

Here we have a complexity-coherence tradeoff, since simple satisficing is also a process with no more than K states, and simple satisficing is more coherent than K-phase satisficing.[16] Agents can opt for greater complexity in the form of K-phase satisficing, or for more coherence in the form of simple satisficing. Why might agents opt for a higher risk of incoherence by switching to K-phase satisficing?

A preliminary reason to do this is that, as we saw, K-phase satisficing maximizes expected utility in the choice-design problem.[17] When agents cannot afford the state complexity of utility maximization, they can still make better expected decisions by shifting

[16] This example shows that K-phase satisficing is vulnerable to a form of incoherence that satisficing does not face. In the other direction, note that any incoherence in satisficing is an incoherence in K-phase satisficing, since satisficing is a type of K-phase satisficing.

[17] To say here that K-phase satisficing maximizes expected utility is to say that it has at least as high expected utility as any other process. It is not to say that K-phase satisficing involves explicitly calculating expected utilities – we have already seen that no K-state process can do this. Thanks to an anonymous referee for pressing me to clarify this point.

from satisficing to K-phase satisficing.

Another reason why agents might adopt K-phase satisficing is suggested by cognate discussions in psychology. A common complaint against simple satisficing is that it exhibits no form of learning. Agents specify a utility threshold in advance and do not change that threshold even after calculating the utilities of several options. To be sure, it is often prohibitively expensive to calculate the utilities of all available options as utility maximization requires. But that does not mean we should allow no learning at all. Many descendants of satisficing allow agents to adjust their utility thresholds through processes of *aspiration adaptation*, learning to set new thresholds based on previously calculated utilities (Selten 1998).

We can think of K-phase satisficing as a computationally restricted form of aspiration adaptation, subject to the constraint that at most K-1 potential adaptations can be made. Aspiration adaptation here involves shifting upwards among the utility thresholds $t_0, \dots, t_{K-1}$. Insofar as many theorists think that aspiration adaptation can often be rational, and insofar as K-phase satisficing represents a feasible way to adapt aspirations with limited computational expense, we will recover further motivation for agents to sometimes make up their minds through K-phase satisficing.

This discussion suggests a more general lesson, since traditional models of aspiration adaptation are also subject to IIA violations for exactly the same reason that K-phase satisficing is.[18] The lesson is that computationally tractable forms of aspiration adaptation during satisficing-style decision making are often good ways to improve decision quality. Although aspiration adaptation may represent a desirable increase in complexity, it often induces a complexity-coherence tradeoff by opening the door to forms of incoherence, such as IIA violations, not present in traditional satisficing procedures. If this is right, then the need for aspiration adaptation can be seen as a second factor driving the complexity-coherence tradeoff. We will also see in Section 6.3 that aspiration adaptation opens a promising new strategy for vindicating the rationality of some troubling ordering effects in cognition.

5 Valence-sensitive inference and informational complexity

Considering more complex forms of information opens new avenues for incoherence. One reason why this happens is that complex information may come demonstrably apart from what agents ultimately care about. When this is so, agents who are sensitive to complex types of information will behave incoherently, which they would not have done if they had ignored complex information. In this section, I expand on this thought to demonstrate a third way in which the complexity-coherence tradeoff can arise.

[18] Roughly, the point is that removing 'aspiration-raising events', such as observing x_1 in our example above, can make previously passed-over list elements, such as x_2, become choiceworthy. This point can be made in more formal detail within most popular models of aspiration adaptation, but it is hard to make the point formally in a way that transcends models.

5.1 The description-experience gap

Information can be provided to agents in two different ways. First, information may be described using verbal or symbolic descriptions. For example, I might tell you the sensitivity of a medical test and the base-rate prevalence of the disease that it tests for. Second, information may be experienced without being described, for example by encountering a mixture of sick and healthy people.

A wave of recent studies has established that agents respond in systematically different ways to information learned through experience rather than through description (Hertwig and Erev 2009; Wulff et al. 2018). In particular, in many contexts agents respond more coherently when information is presented experientially rather than descriptively (Schulze and Hertwig 2021; Wulff et al. 2018). This gap in responding to described versus experienced information is known as the *description-experience gap*.

To see how the description-experience gap bears on the complexity-coherence tradeoff, note that the description-experience gap has been offered as a partial explanation of why nonhuman animals are often more coherent than humans (Hertwig et al. 2018; Schulze and Hertwig 2021). Because humans sometimes learn through description, which raises the risk of incoherent responding, humans are often more incoherent than nonhuman animals, who never learn through description. But note that humans are often faced with the choice of whether and to what extent we will make use of complex descriptive information during decision making. In many such instances, we face a complexity-coherence tradeoff. Making use of complex descriptive information may present a desirable increase in complexity, but it nonetheless comes at the cost of a heightened risk of incoherence.

In this section, I focus on the *informational complexity* of decision making: the amount and complexity of information used during decision making.[19] I show how potentially desirable ways of increasing informational complexity come at the cost of heightened vulnerability to incoherence, generating a complexity-coherence tradeoff for agents who must decide whether to increase the informational complexity of their decision-making processes. I focus on a particular type of informational complexity, the *semantic valence* of descriptions. I show how a range of sophisticated strategies for making use of valenced information can lead to framing effects, while at the same time increasing the expected accuracy of agents' judgments. This raises the possibility of a novel rationalizing explanation for some troubling framing effects, discussed in Section 6.3.

5.2 Attribute framing

Framing effects occur when agents take different attitudes towards equivalent presentations of the same option or decision problem (Bermúdez 2020; Levin et al. 1998; Tversky and Kahneman 1981). For example, we may prefer meat that is 80% lean to meat containing 20% fat, or prefer an 80% chance of a gain to a 20% chance of an equivalent loss.

Framing effects are often regarded as paradigmatic examples of incoherence.[20] Many

[19] Informational complexity is recognized as a type of complexity by many leading taxonomies. For example, Liu and Li (2012) specify ten *complexity contributory factors* of information which increase complexity, Bonner (1994) treats the complexity of informational inputs as one of three types of task complexity, and likewise Wood (1986) treats informational cues as one of three components of task complexity.

[20] For example, Amos Tversky and Daniel Kahneman (1981, p. 453) characterize framing effects as

theorists will be comfortable taking the incoherence of framing effects on board as a plausible observation about coherence. Alternatively, we may support the link between framing and incoherence by showing how framing effects amount to violations of other coherence principles. For example, it is often held as a requirement of coherence that strict preferences be asymmetric:

> **(Asymmetry of Strict Preference)** For all agents S and options o, o' if $o \succ_S o'$ then $o' \not\succ_S o$.

But if I prefer '80% lean' meat to '20% fat' meat, then in many cases there will be some item (turkey, perhaps) such that '80% lean' meat is strictly preferred to turkey, which in turn is strictly preferred to '20% fat' meat. That violates the asymmetry of strict preference, since the same item is both preferred and dispreferred to turkey. Given some structural requirements cases of this form can always be generated from framing effects.[21]

A striking fact about framing effects is that they are much more common in response to descriptive rather than experiential information (Lejarraga and Hertwig 2021).[22] To illustrate why this might be so, consider *attribute framing*. Attribute framing occurs when an attribute of an object or event is manipulated across framings (Levin et al. 1998). For example, agents might prefer meat that is 80% lean to meat containing 20% fat (Levin and Gaeth 1988), or an operation that 60% of patients survive to one which 40% of patients do not survive (Wilson et al. 1987). A bit more carefully: attribute framing involves four elements (Jain et al. 2020). The first three elements are held fixed: a *target entity*, such as ground beef; an *attribute* of the entity, such as fat content; and the *measure* of the attribute, such as 20% fat. What varies across frames is a fourth element, the *semantic valence* of the description used to present the measure of the attribute belonging to the target entity. For example, a single piece of ground beef may be described as having 20% fat (negative semantic valence) or as being 80% lean (positive semantic valence). The entity (ground beef), attribute (fat content) and measure (20% fat) are held fixed.

Attribute framing happens when there is a *valence-consistent shift* in attitudes: agents prefer items whose attributes are framed positively rather than negatively. A primary explanation for this valence-consistent shift in attitudes is that there is an underlying valence-consistent shift in cognitive processing (Levin et al. 1998; Payne et al. 2013). Agents treat valence information as a decision cue by using semantic valence to alter decision-related cognitive processes such as attention, memory and reasoning. For example, agents preferentially attend to positive features of items framed positively and to negative features of items framed negatively (Jain et al. 2020).

It is understandable why agents would treat semantic valence as a decision cue: semantic valence is often correlated with outcome quality. Indeed, agents could do far worse than

violations of 'elementary requirements of consistency and coherence', and Benedetto De Martino and colleagues (De Martino et al. 2006, p. 648) regard framing effects as violations of 'logical consistency across decisions', because they violate extensionality.

[21] For example, it follows from the continuity axiom of von Neumann–Morgenstern theory that some lottery among '20% fat' and '80% lean' meat can take the place of turkey.

[22] Alleged framing effects in response to experiential information (Fu et al. 2018; Gonzalez and Mehlhorn 2016) are rare and sometimes controversial (Kühberger 2021). Precisely for this reason, framing effects are only occasionally documented in infants and nonhumans (Krupenye et al. 2015; Marsh and Kacelnik 2002), and again these effects are controversial (Houston and Wiesner 2020; Kanngiesser and Woike 2016).

to exclusively buy products labeled 'lean' at the grocery store, and the valence-consistent shift in processing improves on this heuristic by allowing other factors to weigh against the impact of a 'lean' label. However, reliance on semantic valence creates the possibility of framing effects, since one and the same object can be described with positive valence or with negative valence without changing any relevant features of the object. And it is just this manipulation in which attribute framing consists.

If feasible strategies for treating semantic valence as a decision cue heighten an agent's risk of incoherent responding, then in deciding whether to incorporate semantic valence agents confront a complexity-coherence tradeoff. Could an increased risk of incoherence be a price worth paying for heightened sensitivity to outcome variation? We will see in Section 6.3 that even many theorists sympathetic to the rationality of framing effects have not wanted to treat paradigmatic cases of attribute framing as rational. However, in the next section, I construct a simple model of a choice situation where the price may be worth paying.

5.3 *Why heed valence?*

I must confess that I often peruse the candy shelf while waiting in the grocery checkout aisle. I quickly scan the available chocolate bars with the goal of purchasing a bar that is high-quality and not too unhealthy. For me, the value of a candy bar increases in its quality q and healthiness h, but decreases with its cost. Let's take a simple model on which value is additive and cost is fixed at 1 util:

$$V(x) = q + h - 1.$$

Let's assume that quality is normally distributed, with mean 0 and variance 3. For simplicity, let's assume that quality and healthiness are uncorrelated, and take healthiness to be a binary variable with equal chance of taking the values -2 (unhealthy) or 0 (healthy).

My perusal of the candy shelf provides me with a noisy signal $\overline{q}$ of candy bar quality. Let's say that:

$$\overline{q} = q + \epsilon.$$

where ϵ is a normally distributed error parameter with mean zero and variance 2, independent of quality and health. When I am in a rush, I make up my mind based only on the quality signal $\overline{q}$. Call this the *quality-only method*. Using the quality-only method, the optimal policy is to purchase a bar just in case $\overline{q} \geq 26/9$, and this policy yields average utility .605 across candy bars.

However, I am a moderately health-conscious chap. I hardly have time to compare nutrition labels, but there are other ways for me to track facts about nutrition. Some candy bars come labeled with words such as 'light', 'diet' or 'skinny'. Let's call such labels 'lean' labels. Let's assume for simplicity that labels are independent of candy bar quality and error signals, and also that labels are 75% reliable indicators of healthiness. More formally, letting Lean be the proposition that a candy bar is labeled 'lean', we will assume that $Pr(h = 0|\textsc{Lean}) = .75$ and $Pr(h = 0|\neg\textsc{Lean}) = .25$.

Suppose I make my decision by combining the quality signal $\overline{q}$ with label information. Call this the *label method*. Now I can do a bit better than before. The optimal policy is to choose bars with a 'lean' label so long as $\overline{q} \geq 13/6$, and bars without a 'lean' label if

$\overline{q} \geq 65/18$. This policy yields average utility .618, an improvement on the quality-only method.

In this model, responding to the semantic valence of descriptions looks like a good way to increase decision quality without spending all day in the checkout line. I will, on scattered occasions, be vulnerable to incoherence. I might pass over a Snickers bar one day, only to buy a Snickers bar the next day after it has been merely relabeled to 'skinny', or more perniciously as '40% lighter than a king size Snickers'. I will pay a quantifiable price in decision quality for my incoherence, but that price is not enough to outweigh the gain in average decision quality from incorporating label information.

Now it might seem that merely relying on the semantic valence of labels could not possibly be a reasonable way to make health-conscious decisions. But in fact, just this one cue takes me a surprisingly long way towards the optimally health-conscious decision policy. Suppose I were to take much longer to make my decision, as a result of which I could deductively determine the true value of h from nutrition labels. Call this the *deductive method*. In this case, the optimal policy would be to choose a bar for which $\overline{q} \geq 13/9$ if it is healthy, or $\overline{q} \geq 13/3$ if it is unhealthy. This policy yields average utility .654.

If I have all day to pick out a candy bar, the deductive method may be worthwhile. But note that the label method of attending only to the semantic valence of labels already realizes 27% of the utility gains reaped by the demanding deductive method. This means that when the deductive method is not feasible or cost-effective, the label method may be a reasonable way for me to make better decisions by incorporating health information into decision making.

The takeaway lesson of this discussion is that in choosing whether to heed or ignore semantic valence in purchasing a candy bar, I confront a complexity-coherence tradeoff. Valence-sensitive decision policies represent a feasible increase in complexity that I may have reason to pursue, even though these policies heighten my risk of incoherent responding. And while I would not dream of telling my readers how to purchase a candy bar, insofar as I am well-described by some model such as the above, I find myself willing to heed valence.

6 Discussion

So far, we have seen evidence for a systematic complexity-coherence tradeoff in cognition. Across a range of cases, feasible increases in the complexity of cognitive processes reduce the expected coherence of the attitudes that result. We explored three of the many factors driving the complexity-coherence tradeoff: procedural complexity (§3), aspiration adaptation (§4) and informational complexity (§5). And we saw how the complexity-coherence tradeoff can be replicated across a variety of coherence requirements, including the transitivity, asymmetry and irreflexivity of strict preference, as well as the requirement to avoid framing effects.

In this section, I discuss normative and descriptive implications of the complexity-coherence tradeoff and survey directions for future research.

6.1 Confronting the complexity-coherence tradeoff

How should agents confront the complexity-coherence tradeoff? The cases in this paper are designed to illustrate why it might sometimes be attractive for agents to privilege complexity over coherence. Making processes more complex is often a good way to increase decision quality at a feasible cost, as in the turn from lexicographic to semilexicographic choice (§3) or an increase in the state-complexity of cognitive processes (§4). And high levels of complexity allow humans to reap the benefits of symbolic knowledge and understanding (§5), which make possible a variety of uniquely human pursuits such as science, mathematics and philosophy. For these reasons, not even the most ardent defender of simple heuristics should deny that more complexity is sometimes better.

However, this is not to say that agents should always privilege complexity over coherence. Traditional discussions in philosophy and cognitive science reveal many reasons that agents may prefer to avoid complex cognitive processes. Complex processes are often slow and cognitively costly. Moreover, it is simply not true that complex processes always outperform simpler processes, even once factors such as time and cognitive costs are ignored (Gigerenzer and Brighton 2009; Wheeler 2020). In this paper, we have enriched the case against complexity by noting another cost of complexity: complexity often comes at the direct expense of coherence.

My aim in this paper is not to suggest that complexity should always take precedence over coherence in cognition. But neither do I want to suggest that coherence should always take precedence over complexity. The complexity-coherence tradeoff, like the accuracy-effort tradeoff, is a genuine tradeoff whose consequences must be carefully measured and weighed. A good way to take the measure and weight of the complexity-coherence tradeoff is to look at how this tradeoff arises in familiar philosophical and scientific debates. I close with a discussion of three applications that may be productive avenues for future research.

6.2 Approximate coherentism

Many theories of rationality hold that unbounded agents are rationally required to be fully coherent. It is tempting to generalize this requirement to cover bounded agents. Although bounded agents may not always be able to achieve full coherence, *approximate coherentists* hold that bounded agents are rationally required to approximate coherence as best they can given their bounds (Staffel 2020; Zynda 1996). For example, Lyle Zynda holds that coherence is an ideal of rationality, and that 'we as epistemic agents ought to approximate this ideal as closely as is possible for us' (Zynda 1996, p. 176). Similarly, Julia Staffel holds that Bayesian norms express 'ideals that imperfect thinkers should approximate' (Staffel 2020, p. 3).[23]

Scientific theories of bounded rationality have typically been suspicious of coherence as a normative standard (Gigerenzer 2019; Gigerenzer and Sturm 2012). This raises a puzzle: what might explain theorists' unique skepticism of coherence as a theory of bounded

[23] Staffel (personal correspondence) notes that the discussion in this section need not threaten her view that an agent's credences are more propositionally rational the more closely they approximate the ideal credence function. See also Thorstad (forthcoming) for discussion of the compatibility of coherentist norms on attitudes with non-coherentist norms on processes.

rationality against the background of widespread support for coherence requirements on unbounded agents?

One natural answer is that bounded agents confront tradeoffs that unbounded agents may avoid. For example, Thorstad (forthcoming) argues that bounded agents often face a systematic accuracy-coherence tradeoff in cognition. That is, they must choose between a range of feasible strategies, where the strategies that produce, in expectation, the most coherent results differ from those that produce, in expectation, the most accurate results. Thorstad suggests that agents may sometimes be rationally permitted, or even required, to opt for less coherent strategies in order to promote the formation of accurate beliefs.

This paper extends Thorstad's suggestion by illustrating another systematic tradeoff in cognition: the complexity-coherence tradeoff.[24] Just as it is natural to take the accuracy-coherence tradeoff to suggest that rational agents sometimes sacrifice a degree of coherence to gain in accuracy, so too it is natural to take the complexity-coherence tradeoff to suggest that rational agents sometimes sacrifice a degree of coherence to gain the myriad benefits of cognitive complexity. If this is right, then it lends support and robustness to a tradeoff-based explanation of the difficulties facing approximate coherentism.

It is, of course, open to approximate coherentists to argue that there is a special type of rationality which is, by nature, exhausted by coherence. This would insulate approximate coherentists from tradeoff-based criticism, but this move has important theoretical costs. For one thing, leading approximate coherentists have justified their view by arguing that approximate coherence tracks other valuable cognitive goals, such as accuracy (De Bona and Staffel 2018; Staffel 2020). These arguments will be lost on a picture on which coherence trades off against accuracy, complexity and other desiderata.

Moreover, once tradeoffs are admitted, approximate coherentists will struggle to recover important normative data that are often invoked in theorizing about rationality. One such datum is the authority of rationality: rationality is authoritative over behavior (Kauppinen 2021; Kiesewetter 2017). When coherence comes at the expense of other desirable properties such as accuracy or the benefits of complex cognition, it is increasingly difficult to see how approximate coherentism could be authoritative. In Sections 3-5 of this paper, we saw a number of cases in which a small decrease in expected coherence could provide important benefits, including provable gains in expected utility. I suggested in Section 6.1 that in many such cases, it is appropriate to accept a small decrease in expected coherence in exchange for the benefits of complex cognition. But this suggestion is incompatible with the authority of rationality unless we reject approximate coherentism as an account of rationality.

Another datum is the value of rationality: rationality has significant value (Horowitz 2014; Wedgwood 2017). As coherence begins to conflict with other valuable ends such as accuracy, utility, or the myriad other benefits of complex cognition, it begins to look better in some cases to be incoherent than to be coherent. On an approximate coherentist view of rationality, this amounts to the claim that it would be better to be irrational

[24] Note that the complexity-coherence tradeoff may come apart from the accuracy-coherence tradeoff. To see this, consider all the cases in which less-is-more effects obtain (Geman et al. 1992; Gigerenzer and Brighton 2009): that is, in which increasing complexity tends to decrease accuracy. In all such cases where the complexity-coherence tradeoff continues to hold, the complexity-coherence tradeoff will come apart from the accuracy-coherence tradeoff, which no longer holds. Thanks to an anonymous referee for pressing me to address this point.

than to be rational. On one common reading of the datum, rationality cannot be less valuable than irrationality. This reading would immediately falsify the datum in the presence of tradeoffs. Perhaps weaker readings of the value of rationality could survive the challenges raised in this paper, but this is far from a sure thing, and this weaker reading would conflict with leading argumentative uses of the value of rationality (Horowitz and Dogramaci 2016; Steglich-Petersen 2011).

Traditional objections to coherence-based theories of rationality have held that coherence is an epiphenomenon, in the sense that coherence requirements are not fundamental requirements of rationality, but rather fall out as a consequence of other rational requirements (Kolodny 2005). For example, on evidentialist theories of rational belief, beliefs are rational only if they are evidentially supported. On many views, evidence cannot support an incoherent combination of beliefs. This means that the requirement to hold coherent beliefs need not be taken as a primitive rational requirement. After all, this requirement comes for free given evidentialism.

Typically, the epiphenomenal objection to coherence-based theories of rationality is not used to show that coherence requirements are false, but only that they are normatively non-fundamental. However, for bounded agents, tradeoffs emerge that spoil the coincidence between coherence and other cognitive goals. Now we cannot simply hold that coherentism, even in its approximate form, is entailed by theories such as accuracy maximization, utility maximization, or other theories which seek the benefits of complex cognition. When forced to choose between coherence and other cognitive goals, many agents will often choose to sacrifice some degree of coherence. This suggests that maximizing coherence was never, in itself, a fundamental normative requirement, but only, as the epiphenomenalist suggests, a condition that agents were content to satisfy when it did not come at the expense of other goals.

While the arguments in this paper have been focused on approximate coherentism, it may be worth exploring whether generalizations of these arguments could put pressure on other one-factor theories, in which a single goal such as accuracy is to be maximized at all costs. Just as it may look problematically tradeoff-insensitive to always choose coherence over competing cognitive goals such as accuracy and utility, so too it may look problematically tradeoff-insensitive to always choose accuracy over competing goals such as utility, coherence, or effort minimization. If this is right, then we may want to update in favor of tradeoff-sensitive theories such as instrumentalism (Steglich-Petersen 2011) and consequentialism (Stich 1990) which treat various goals such as coherence, accuracy or the myriad benefits of complex cognition as competing goals and give explicit accounts of how these competing goals are to be traded off during cognition.

6.3 *Vindicatory epistemology*

In the twentieth century, scientific theorizing about rationality swung from a mid-century optimism that regarded most humans as highly rational most of the time to a late-century period of pessimism in which human rationality was regarded much more dimly.[25] This century has seen the emergence of a newly empiricized optimism that once again regards most humans as highly rational most of the time (Gigerenzer and Selten 2001; Lieder and Griffiths 2020).

[25] See Samuels et al. (2002); Sturm (2012) and Thorstad (forthcoming b) for discussion.

The foundation for optimism is the program of vindicatory epistemology (Thorstad forthcoming b), which aims to recast seemingly irrational cognitions as fully rational. Within philosophy, phenomena as diverse as framing effects (Bermúdez 2020), polarization (Dorst 2023), randomization (Icard 2021), and attention to sunk costs (Kelly 2004) have been given rationalizing reconstructions.

An under-utilized strategy in many recent vindicatory arguments is the appeal to tradeoffs. To demonstrate the strength of this strategy, consider attribute framing (§5.2). Focus in particular on an agent who prefers '80% lean' beef to '20% fat' beef. Although some theorists have recently aimed to vindicate the rationality of framing effects, most theorists have struggled to vindicate attribute framing. For example, a recent book-length defense of the rationality of some framing effects by José Luis Bermúdez (2020) directly argues that the preference for '80% lean' over '20% fat' beef is irrational: after all, Bermúdez notes, agents would be disposed to withdraw the preference on realizing that '80% lean' and '20% fat' beef are the same thing.

The complexity-coherence tradeoff provides a defense of attribute framing that is fully consistent with Bermudez's observation that attribute framing would be no longer rational once recognized by the agent.[26] Bounded agents may rationally rely on label valence as a quick and reasonably effective indicator of quality. They do this because the process of judging quality by label valence is quick and fairly reliable, and because it is not worth investing more cognitive resources into ordinary supermarket purchases. This rationalizes the use of processes which produce attribute framing effects, but it does not rationalize the continued presence of framing effects after a more complex process which detects the equivalence of '80% lean' and '20% fat' beef. The strategy of ignoring detected equivalences is still quick, but no longer reliable. In this way, the tradeoff-based vindication of attribute framing helps us to recover a result that even many vindicatory theorists have struggled to deliver: processes that lead to attribute framing may be fully rational, even though it would not be rational to retain a preference for '80% lean' beef over '20% fat' beef upon discovering they are equivalent.

More generally, it may be worth exploring whether the complexity-coherence tradeoff can ground other vindicatory results that were difficult for previous accounts to deliver. We have already seen several such results throughout this paper. Section 3 gave a new vindication of vulnerability to intransitive preferences by casting this vulnerability as the result of potentially desirable increases in procedural complexity, as in the turn from tallying to near-tallying. This complements existing attempts to vindicate intransitive preference (Houston et al. 2007; Mandler 2005), which have met with opposition. Section 4 gave a new vindication of vulnerability to ordering effects by suggesting that within a range of feasible strategies, increases in the state complexity of cognitive processes may be expected utility maximizing, even as they increase the risk of violating the principle of independence of irrelevant alternatives.

One noteworthy application of the complexity-coherence tradeoff comes in defending the rationality of heuristic strategies against the charge that heuristics are incoherent. It is often thought that fast and frugal heuristic strategies can be rational in environments where they are accurate and cognitively efficient (Johnson and Payne 1985), or where they reduce the risk of overfitting decision processes to sparse data (Gigerenzer and Brighton 2009). Nevertheless, opponents counter that heuristic cognition is in an important sense

[26] For a related argument, see Sher and McKenzie (2006).

irrational, because heuristics occasionally return incoherent judgments. The complexity-coherence tradeoff raises the possibility of meeting this challenge on its own turf. Even if we accept coherence as a normative standard, it does not follow from the fact that heuristics sometimes produce incoherent attitudes that heuristics are irrational. After all, the proposal is to replace heuristic strategies with more complex nonheuristic strategies. Insofar as there is often a tradeoff between complexity and coherence in cognition, we should not accept without an argument that these complex replacements will lower rather than raise an agent's vulnerability to incoherence. If that is right, then even approximate coherentists may often treat heuristic cognition as rationally obligatory, rather than irrational. I expand on this point in Section 6.4.

Finally, the discussion in this paper reminds us of the vindicatory potential of attending to presentation formats.[27] Ralph Hertwig and colleagues have noted a marked shift in experimental paradigms from mid-century paradigms which often presented information experientially to more recent paradigms which more often present information descriptively (Hertwig and Erev 2009; Schulze and Hertwig 2021; Wulff et al. 2018). Hertwig and colleagues have noted that this change in experimental paradigms coincides with a shift towards decreased rational performance on tasks and increased skepticism by theorists about the rationality of human judgment and decision making. Hertwig and colleagues suggest that an important vindicatory strategy involves careful scrutiny of the way in which information is presented to agents and how performance may be improved through more helpful ways of presenting information. This insight is supported by previous work in other paradigms, such as the widely replicated funding that agents incorporate base rates significantly better when information is presented using environmental frequencies rather than statistical descriptions (Gigerenzer and Hoffrage 1995). Although the description-experience gap has occasioned a rise in attention among cognitive scientists to the vindicatory potential of the format in which information is presented to experimental subjects, recent vindicatory work in philosophy has not often drawn on the rational importance of presentation formats. It would be productive for future philosophical work to further explore the vindicatory potential of presentation formats.

6.4 Dual process theories of cognition

Dual process theories are among the best-known and most controversial (Keren and Schul 2009; Melnikoff and Bargh 2018) approaches in cognitive science today. Dual process theories claim that humans possess two types of cognitive processes, which can be driven apart not only in their evolutionary history, but also in their performance along a number of dimensions. For example, Jonathan Evans and Keith Stanovich (2013) list a number of typical correlates of Type 1 and Type 2 processes (Table 2).

A growing voice of protest against dual process theories argues that these correlations are far from typical: they break, not just on occasion, but often quite systematically. For example, defenders of fast-and-frugal heuristics argue that Type 1 processes are rule-based rather than associative, since they rely on precisely specifiable rule-based heuristics

[27] To be clear, what the paper provides is an illustration of the vindicatory potential of attention to presentation formats. This potential is also illustrated by other recent results, giving increased support to attending to presentation formats as a vindicatory strategy. Thanks to a referee for pressing me to address this point.

Type 1 Processes	Type 2 Processes
Fast	Slow
High capacity	Capacity limited
Parallel	Serial
Nonconscious	Conscious
Biased responses	Normative responses
Contextualized	Abstract
Automatic	Controlled
Associative	Rule-based
Experience-based decision making	Consequential decision making
Independent of cognitive ability	Correlated with cognitive ability

Table 2: Typical correlates of Type 1 and Type 2 processes, Evans and Stanovich (2013).

(Gigerenzer 2011). One such objection can be survived, but if many of the alleged typical correlations were to systematically break, dual process theories would face two types of pressure. First, it would become increasingly attractive to take dual process theories to be ill-defined as typical correlations used to introduce the distinction between Type 1 and Type 2 processes began to fall in succession. Second, we would increasingly lose the claimed explanatory payoffs of dual process accounts, since many explanations in this tradition use typical correlates as evidence for attributing processes, which can then be used to explain behavior. However, this strategy would not work if the typical correlates did not correlate in the intended way.

Wim de Neys and colleagues have recently argued that another alleged correlation breaks systematically (Bago and De Neys 2017; De Neys forthcoming). Namely, they have argued that in a wide range of cases, Type 1 processes are slow and Type 2 processes are fast. This combines with earlier charges to put pressure on the well-definedness and explanatory application of dual process theories. Although de Neys' findings question one of the most central and defining correlations in dual process theory, perhaps the charge is survivable: de Neys himself ultimately recommends adopting a new version of dual process theory which scrubs even the dichotomy between fast and slow cognition from the list of typical correlates (De Neys forthcoming).

The complexity-coherence tradeoff casts doubt on another correlation alleged by dual process theories: that Type 1 processes are biased and Type 2 processes are normative (Epstein 1994; Evans and Stanovich 2013; Kahneman and Frederick 2002). Although there are some readings of bias on which this correlation may hold, many dual process theorists hold and apply a normative theory closely tied to requirements of coherence or structural rationality (Evans 2008; Evans and Stanovich 2013). Indeed, many biases are simply defined as deviations from the preference axioms, probability axioms, and other requirements of structural rationality (Gilovich et al. 2002; Kahneman et al. 1982). On this reading, it is simply not true to say that Type 2 processes, showing signs of complexity such as consciousness, abstraction, capacity limits and correlation with cognitive ability, are disposed to produce more normative responses than Type 1 processes. If there is indeed a systematic tradeoff between complexity and coherence in cognition, then we should expect Type 1 processes to produce more coherent responses than Type 2 processes do. On a broadly coherence-based theory, this means we should expect Type 1 processes to produce more normative responses than Type 2 processes do.

It is, of course, open to dual process theorists to scrub another central correlation from their list: that Type 2 processes are normative in a coherence-based sense, and Type 1 processes are biased and non-normative in the same sense. This would also involve withdrawing many explanatory applications of dual process theory, including a great number of recent normative arguments. However, no theory can survive indefinite reformulation, and with each abandoned dichotomy there is increasing plausibility to the critics' suggestion that dual process theories do not pick out a genuine distinction between existing psychological processes.

What would be lost if dual process theory were to fall? For one thing, dual process theory has become a central explanatory framework within many domains of psychology, including social learning (Smith and DeCoster 2000), judgment and decision making (Kahneman 2011), the psychology of reasoning (Evans 2011) and metacognition (Thompson 2009). For example, within judgment and decision making, human incoherence is often explained as the result of Type 1 processing (Kahneman 2011). But if complexity emerges as a significant source of incoherence, then it is no longer so obvious that incoherence should be explained by the fact that agents used simple, Type 1 processes. Other descriptive applications of dual process theory may come under similar pressure.

The fall of dual process theorizing would also challenge some claimed normative implications of the program, opening new avenues for vindicatory epistemology. At least two implications bear emphasis. First, dual process theorizing has been used in debunking explanations meant to demonstrate the irrationality of disputed judgments, including cognitive biases (Kahneman 2011) and nonconsequentialist moral intuitions (Greene et al. 2008; Haidt 2001). These judgments are claimed to be the results of unreliable Type 1 processes, and therefore it is suggested that the judgments themselves should be suspect. However, outside of a dual process framework, these complaints could not get traction without substantial reformulation, and it is far from clear that salvage claims would be normatively or descriptively persuasive.

Second, nudging theorists have argued that Type 1 processes lead to systematically irrational behaviors and have sought to design interventions that co-opt irrational Type 1 processes to produce better outcomes (Sunstein 2014; Thaler and Sunstein 2008).[28] Opponents of nudging have questioned the assumptions that nudging targets irrational behaviors produced by Type 1 processes, and that nudging works through co-opting Type 1 processes (Bovens 2009; Grüne-Yanoff 2012). Instead, they have proposed a program of *boosting* which aims to improve decision outcomes by enriching cognitive environments or providing agents with useful cognitive tools, without any assumption of irrationality or underlying distinction between two types of cognitive processes (Grüne-Yanoff and Hertwig 2016). Challenges to dual process theory are, by extension, arguments for boosting instead of nudging, both as a theoretical framework and as a set of policy interventions.

[28] See (Grüne-Yanoff and Hertwig 2016) for discussion of the relationship between nudge theory and dual process theory, among other components of the heuristics and biases program.

6.5 Concluding thoughts

In this paper, we have seen evidence for a systematic complexity-coherence tradeoff in cognition and seen how thinking through the complexity-coherence tradeoff may shed useful light on existing philosophical and scientific debates. The proof is, as they say, in the pudding, and it is in wading through the situational implications of the complexity-coherence tradeoff that we will get a better handle on the nature and extent of the tradeoff, as well as on what the complexity-coherence tradeoff might imply for the study of human cognition.

References

Arkes, Hal and Ayton, Peter 1999, 'The sunk cost and Concorde effects: Are humans less rational than lower animals?' *Psychological Bulletin* 125:591–600.

Bago, Bence and De Neys, Wim 2017, 'Fast logic? Examining the time course assumption of dual process theory', *Cognition* 158:90–109.

Bennett, Charles 1988, 'Logical depth and physical complexity', in Herken, Rolf (ed.), *The universal Turing machine: A half-century survey* (Oxford: Oxford University Press), 227–57.

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

Bonner, Sarah 1994, 'A model of the effects of audit task complexity', *Accounting, Organizations and Society* 19:213–34.

Bossaerts, Peter and Murawski, Carsten 2017, 'Computational complexity and human decision-making', *Trends in Cognitive Sciences* 21:917–29.

Bovens, Luc 2009, 'The ethics of Nudge', in Grüene-Yanoff, Till and Hansson, Sven Ove (eds.), *Preference change* (Springer), 207–19.

Crutchfield, James and Young, Karl 1989, 'Inferring statistical complexity', *Physical Review Letters* 63:105–8.

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.

Dawkins, Richard and Brockmann, Jane 1980, 'Do digger wasps commit the Concorde fallacy?' *Animal Behavior* 28:892–6.

De Bona, Glauber and Staffel, Julia 2018, 'Why be (approximately) coherent?' *Analysis* 78:405–15.

De Martino, Benedetto, Kumaran, Dharshan, Seymour, Ben and Dolan, Raymond 2006, 'Frames, biases, and rational decision-making in the human brain', *Science* 313:684–7.

De Neys, Wim forthcoming, 'Advancing theorizing about fast-and-slow thinking', *Behavioral and Brain Sciences* forthcoming.

Dorst, Kevin 2023, 'Rational polarization', *Philosophical Review* 132:355–458.

Epstein, Seymour 1994, 'Integration of the cognitive and the psychodynamic unconscious', *American Psychologist* 49:709–24.

Evans, Jonathan 2008, 'Dual-processing accounts of reasoning, judgment, and social cognition', *Annual Review of Psychology* 59:225–78.

Evans, Jonathan 2011, 'Dual-process theories of reasoning: Contemporary issues and developmental applications', *Developmental Review* 31:86–102.

Evans, Jonathan and Stanovich, Keith 2013, 'Dual-process theories of higher cognition: Advancing the debate', *Perspectives on Psychological Science* 8:223–41.

Fu, Lisha, Yu, Junjie, Ni, Shiguang and Li, Hong 2018, 'Reduced framing effect: Experience adjusts affective forecasting with losses', *Journal of Experimental Social Psychology* 76:231–8.

Gell-Mann, Murray 1995, 'What is complexity?' *Complexity* 1:16–19.

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

Gigerenzer, Gerd 2011, 'Personal reflections on theory and psychology', *Theory and Psychology* 20:733–43.

Gigerenzer, Gerd 2019, 'Axiomatic rationality and ecological rationality', *Synthese* 194:3547–64.

Gigerenzer, Gerd and Brighton, Henry 2009, 'Homo heuristicus: Why biased minds make better inferences', *Topics in Cognitive Science* 1:107–43.

Gigerenzer, Gerd and Hoffrage, Ulrich 1995, 'How to improve Bayesian reasoning without instruction: Frequency formats', *Psychological Review* 102:684–704.

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

Gigerenzer, Gerd and Sturm, Thomas 2012, 'How (far) can rationality be naturalized?' *Synthese* 187:243–68.

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

Gonzalez, Cleotilde and Mehlhorn, Katja 2016, 'Framing from experience: Cognitive processes and predictions of risky choice', *Cognitive Science* 40:1163–91.

Gopnik, Alison 2014, 'The surprising probability gurus wearing diapers', *The Wall Street Journal* January 10, 2014.

Greene, Joshua, Morellia, Sylvia, Lowenberg, Kelly, Nystrom, Leigh and Cohen, Jonathan 2008, 'Cognitive load selectively interferes with utilitarian moral judgment', *Cognition* 107:1144–54.

Grüne-Yanoff, Till 2012, 'Old wine in new casks: Libertarian paternalism still violates liberal principles', *Social Choice and Welfare* 38:635–45.

Grüne-Yanoff, Till and Hertwig, Ralph 2016, 'Nudge versus boost: How coherent are policy and theory?', *Minds and Machines* 26:149–83.

Haidt, Jonathan 2001, 'The emotional dog and its rational tail: A social intuitionist approach to moral judgment', *Psychological Review* 108:814–34.

Heitz, Richard 2014, 'The speed-accuracy tradeoff: History, physiology, methodology, and behavior', *Frontiers in Neuroscience* 8:1–19.

Hertwig, Ralph and Erev, Ido 2009, 'The description-experience gap in risky choice', *Trends in Cognitive Sciences* 13:517–23.

Hertwig, Ralph, Hogarth, Robin and Lejarraga, Tomás 2018, 'Experience and description: Exploring two paths to knowledge', *Current Directions in Psychological Science* 27:123–8.

Horowitz, Sophie 2014, 'Immoderately rational', *Philosophical Studies* 167:41–56.

Horowitz, Sophie and Dogramaci, Sinan 2016, 'An argument for uniqueness about evidential support', *Philosophical Issues* 26:130–47.

Houston, Alasdair, McNamara, John and Steer, Mark 2007, 'Violations of transitivity under fitness maximization', *Biology Letters* 3:365–7.

Houston, Alasdair and Wiesner, Karoline 2020, 'Gains v. losses, or context dependence generated by confusion?', *Animal Cognition* 23:361–6.

Icard, Thomas 2021, 'Why be random?' *Mind* 130:111–39.

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.

Kahneman, Daniel 2011, *Thinking, fast and slow* (Farrar, Straus and Giroux).

Kahneman, Daniel and Frederick, Shane 2002, 'Representativeness revisited: Attribute substitution in intuitive judgment', in Gilovich, Thomas; Griffin, Dale; and Kahneman, Daniel (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 48–81 (Cambridge University Press).

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

Kanngiesser, Patricia and Woike, Jan 2016, 'Framing the debate on human-like framing effects in bonobos and chimpanzees: A comment on Krupenye et al (2015)', *Biology Letters* 12:20150959.

Kauppinen, Antti 2021, 'Rationality as the rule of reason', *Nôus* 55:538–59.

Kelly, Thomas 2004, 'Sunk costs, rationality, and acting for the sake of the past', *Nôus* 38:60–85.

Keren, Gideon and Schul, Yaacov 2009, 'Two is not always better than one: A critical evaluation of two-system theories', *Perspectives on Psychological Science* 4:533–50.

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

Kolmogorov, Andrey 1965, 'Three approaches to the quantitative definition of information', *Problems of Information Transmission* 1:1–17.

Kolodny, Niko 2005, 'Why be rational?' *Mind* 114:509–63.

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

Kühberger, Anton 2021, 'Risky choice framing by experience: A methodological note', *Judgment and Decision Making* 16:1314–23.

Ladyman, James and Wiesner, Karoline 2020, *What is a complex system?* (Yale University Press).

Lejarraga, Tomás and Hertwig, Ralph 2021, 'How experimental methods shaped views on human competence and rationality', *Psychological Bulletin* 147:535–64.

Levin, Irwin and Gaeth, Gary 1988, 'How consumers are affected by the framing of attribute information before and after consuming the product', *Journal of Consumer Research* 15:374–8.

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* 76:149–88.

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.

Liu, Peng and Li, Zhizhong 2012, 'Task complexity: A review and conceptualization frmaework', *International Journal of Industrial Ergonomics* 42: 553-68.

Mandler, Michael 2025, 'Incomplete preferences and rational intransitivity of choice', *Games and Economic Behavior* 50: 225-77.

Marsh, Barnaby and Kacelnik, Alex 2002, 'Framing effects and risky decisions in starlings', *Proceedings of the National Academy of Sciences* 99: 3552-5.

Melnikoff, David and Bargh, John 2018, 'The mythical number two,' *Trends in Cognitive Sciences* 22: 280-93.

Morton, Adam 2010, 'Human bounds: Rationality for our species', *Synthese* 176: 5-21.

Okasha, Samir 2018, *Agents and goals in evolution* (Oxford University Press).

Oprea, Ryan 2020, 'What makes a rule complex?', *American Economic Review* 110: 3913-51.

Parfit, Derek, *Reasons and persons* (Oxford University Press).

Payne, John, Sagra, Namika, Shu, Suzanne, Appelt, Kristin and Johnson, Eric 2013, 'Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect', *Journal of Risk and Uncertainty* 46: 27-50.

Railton, Peter 1984, 'Alienation, consequentialism, and the demands of morality', *Philosophy and Public Affairs* 13: 134-71.

Rubinstein, Ariel 1986, 'Finite automata play repeated prisoner's dilemma', *Journal of Economic Theory* 39: 83-96.

Salant, Yuval 2011, 'Procedural analysis of choice rules with applications to bounded ratoinality', *American Economic Review* 101: 724-48.

Samuels, Richard and Stich, Stephen and Bishop, Michael 202, 'Ending the rationality wars: How to make disputes about human rationality disappear', in *Common sense, reasoning and rationality* ed. Renee Elio (Oxford University Press): 236-68.

Schmid, Bernhard 2016, 'Decision-making: Are plants more rational than animals?', *Current Biology* 26: R675-8.

Schuck-Paim, Cynthia 2022, 'Rationality in risk-sensitive foraging choices by starlings', *Animal Behavior* 64: 869-79.

Schulze, Christin and Hertwig, Ralph 2021, 'A description-experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups', *Cognition* 210: 104580.

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

Selten, Reinhard 1988, 'Aspiration adaptation theory', *Journal of Mathematical Psychology* 42: 191-214.

Shafir, Eldar and LeBoeuf, Robyn 2002, 'Rationality', *Annual Review of Psychology* 53:491–517.

Shafir, Sharoni 1994, 'Intransitivity of preferences in honey bees: Support for "comparative" evaluation of foraging options', *Animal Behavior* 48:55–67.

Shannon, Claude 1948, 'A mathematical theory of communication', *The Bell System Technical Journal* 27:379–423.

Sher, Shlomi and McKenzie, Craig 2006, 'Information leakage from logically equivalent frames', *Cognition* 101:467–94.

Smith, Eliot and DeCoster, Jamie 2000, 'Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems', *Psychology and Social Psychology Review* 4:108–31.

Spurrett, David 2021, 'The descent of preferences', *British Journal for the Philosophy of Science* 72:485–510.

Staffel, Julia 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.

Steglich-Petersen, Asbjørn 2011, 'How to be a teleologist about epistemic reasons', in Reisner, Andreas and Steglich-Petersen, Asbjørn (eds.), *Reasons for belief*, 13–33 (Cambridge University Press).

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

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

Sunstein, Cass 2014, *Why nudge? The politics of libertarian paternalism* (Yale University Press).

Sweis, Brian, Abram, Samantha, Schmidt, Brandy, Seeland, Kelsey, MacDonald, Angus, Thomas, Mark and Redish, A. David 2018, 'Sensitivity to "sunk costs" in mice, rats and humans', *Science* 361:178–81.

Thaler, Richard and Sunstein, Cass 2008, *Nudge: Improving decisions about health, wealth, and happiness* (Yale University Press).

Thompson, Valerie 2009, 'Dual-process theories: A metacognitive perspective', in Evans, Jonathan and Frankish, Keith (eds.), *In two minds: Dual processes and beyond* (Oxford University Press).

Thorstad, David forthcoming, 'The accuracy-coherence tradeoff in cognition', *British Journal for the Philosophy of Science* forthcoming.

— forthcoming b, *Inquiry under bounds* (Oxford University Press).

Todd, Peter and Gigerenzer, Gerd 2012, 'What is ecological rationality?' in Todd, Peter and Gigerenzer, Gerd (eds.), *Ecological rationality: Intelligence in the world*, 3–30 (Oxford University Press).

Tversky, Amos and Kahneman, Daniel 1981, 'The framing of decisions and the psychology of choice', *Science* 211:453–8.

Wedgwood, Ralph 2017, *The value of rationality* (Oxford University Press).

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

Wilson, Dawn, Kaplan, Robert and Schneiderman, Lawrence 1987, 'Framing of decisions and selections of alternatives in health care', *Social Behaviour* 2:51–9.

Wood, Robert 1986, 'Task complexity: Definition of the construct', *Organizational Behavior and Human Decision Processes* 37:60–82.

Wulff, Dirk, Mergenthaler-Canseco, Max and Hertwig, Ralph 2018, 'A meta-analytic review of two modes of learning and the description-experience gap', *Psychological Bulletin* 144:140–76.

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