The complexity-coherence tradeoff in cognition

Abstract

I argue that bounded agents face a systematic complexity-coherence tradeoff in cognition. Agents must choose whether to structure their cognition in more complex ways, or in ways more likely to promote coherence. I illustrate the complexity-coherence tradeoff by examining three types of complexity: procedural complexity, informational complexity, and state complexity. In each case, I show how feasible strategies for increasing complexity along the relevant dimension often come at the expense of a heightened vulnerability to incoherence. I discuss normative and descriptive implications of the complexity-coherence tradeoff, including a novel challenge to coherence-based theories of bounded rationality, renewed support for the rationality of heuristic cognition, and a deepening of traditional challenges to dual-process theories of cognition.

1 Introduction

Here is a puzzling fact. 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 (Gilovich et al. 2002; 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.[1] In the limiting case of plant cognition, no incoherence has ever been observed (Schmid 2016). Why would the most complex creatures on earth also be the least coherent?

This 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 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 that the inverse relationship between complexity and coherence is not a coincidence, but rather part of a systematic complexity-coherence

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

tradeoff in cognition (Stanovich 2013; Thorstad forthcoming a).[2] In many cases, cognitive complexity comes at the cost of increased vulnerability to incoherence.

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 (Section 2). Second, I argue that the complexity-coherence tradeoff often obtains and catalog three of the factors driving the complexity-coherence tradeoff: procedural complexity (Section 3), aspiration adaptation (Section 4) and informational complexity (Section 5). Finally, I draw out normative and descriptive implications of the complexity-coherence tradeoff and sketch directions for future work (Section 6).

2 Clarifying the target

What does it mean to speak of a complexity-coherence tradeoff in cognition? We can get a handle on what this means by thinking about another famous tradeoff in cognition: the accuracy-effort tradeoff (Johnson and Payne 1985). It is commonly noted that there is a systematic tradeoff between the effort expended by cognitive processes and the accuracy of the beliefs that result. In expectation, putting more effort into deliberation often produces more accurate beliefs. Thinking of the complexity-coherence tradeoff as analogous to the accuracy-effort tradeoff suggests five clarifications that will help us to understand the complexity-coherence tradeoff in cognition.

First, the complexity-coherence tradeoff, like the accuracy-effort tradeoff, occurs at the level of cognitive processes rather than the attitudes they produce. To speak of an accuracy-effort tradeoff in cognition is to say that agents must choose among a range of feasible cognitive processes and that the processes which produce, in expectation, the most accurate judgments are often distinct from the processes which expend, in expectation, the least effort. Similarly, to speak of a complexity-coherence tradeoff in cognition is to say that agents must choose among a range of feasible cognitive processes and that the most complex of these processes are not always, in expectation, the processes which produce the most coherent attitudes. 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.[3] But my interest in this paper is only with the cognitive processes that agents select during their lifetimes.

Second, the complexity-coherence tradeoff, like the accuracy-effort tradeoff, occurs often but not always. In some situations, the accuracy-effort tradeoff is non-existent or even reversed (Geman et al. 1992; Gigerenzer and Brighton 2009; Wheeler 2020). For this reason, research on the accuracy-effort tradeoff has concentrated on identifying the factors which drive the presence or absence of this tradeoff. Similarly, my claim is that complexity and coherence often trade off, not that they always trade off. 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.

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

[3] Indeed, Okasha (2018) and Spurrett (2021) argue that the evolutionary factors favoring complex cognition do not always favor coherence.

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

Fourth, the notion of coherence is notoriously fraught. One problem is that the notion of coherence is picked out by a patchwork of at most roughly coextensive terms including consistency, coherence, structural rationality (Worsnip 2021), axiomatic rationality (Gigerenzer 2019), and the Standard Picture (Stein 1996; Thorstad forthcoming c). Another problem is that theorists disagree amongst themselves about the extension of each of these terms. For the most part, I will address this problem by focusing on simple forms of incoherence widely accepted across theories and views, such as intransitivity, symmetry and reflexivity of strict preference. 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).[4] 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 et al. 2013; 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; McAllister 2003), 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 (Section 3), state complexity (Section 4), and informational complexity (Section 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.

[4] Some approaches diverge even further from these. For example, theorists have noted a tradeoff between complexity and coherence in landscape design (Kaplan and Kaplan 1989).

3 Lexicographic choice and procedural complexity

Our first example of a complexity-coherence tradeoff is due to David Thorstad (forthcoming a). Suppose you are buying a car. One way you could choose between cars is through *lexicographic choice* (Fishburn 1974). You would rank the features of cars by importance. Perhaps the most important feature is an automatic transmission; then price; then safety rating. You then select the car performing best on the most important feature: having an automatic transmission. If several cars perform best on this feature, the remaining cars are ranked by their performance on the second feature: price. Choice proceeds in this way until a single option remains, then this option is chosen.

More formally, lexicographic choosers face a set $O = \{o_1, \ldots, o_m\}$ of options, in this case cars. They compare cars using some features $f_1, \ldots, f_n$ such as transmission type, price and safety rating, ranked by descending importance. For each feature f_i, they settle on a value function V_i ranking the goodness of each value that f_i can take. For example, perhaps $V_1(x) = 1$ if x is 'automatic' and 0 otherwise. If some option maximizes V_1, that option is chosen. Otherwise, the options maximizing V_1 are compared according to V_2, repeating until one option remains.

Lexicographic choice is rarely a good way to buy a car. Our lexicographic chooser always buys the cheapest automatic on the market, ignoring its other features. A popular way to improve on lexicographic choice is semilexicographic choice (Tversky 1969). Semilexicographic choosers identify, for each feature f_i, a *just noticeable difference* in value to overlook. Perhaps they will ignore price differences under one thousand dollars and differences in safety rating of no more than a star. Choice continues as before, except that options are no longer eliminated if they fall within a just noticeable difference of the leading option. Our semilexicographic chooser may no longer buy the cheapest automatic car on the market — price differences under a thousand dollars can be compensated for by increased safety.

But semilexicographic choice is less coherent than lexicographic choice.[5] 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 $o' \succ o''$ then $o \succ o''$.

Lexicographic choosers always satisfy transitivity, but semilexicographic choosers may not. Suppose our semilexicographic chooser is confronted with the following three automatic cars:

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

[5]This example shows that semilexicographic choice is vulnerable to one type of incoherence which does not affect lexicographic choice. In the other direction, every incoherence in lexicographic choice is an incoherence in semilexicographic choice, since lexicographic choice is a special case of semilexicographic choice.

Between Car A and Car B, she will select Car A. Between Car B and Car C, she will select car B. Between Car C and Car A, she will select Car C. It is natural to interpret this as a pattern of intransitive strict preferences.[6]

Thorstad (forthcoming a) suggests that the reason why semilexicographic choice is less coherent than lexicographic choice lies in its added complexity. Semilexicographic choice adds an extra tie-breaking step that is selectively used to settle close calls. In this case, safety ratings are used to break the ties between Car A and Car B, as well as between Car B and C, but not between Car A and Car C. Although this tie-breaking step may improve average decision quality, it creates a risk of incoherence because the tie-breakers are invoked to settle only some, but not all of the binary choices between cars. If that is right, then in choosing between lexicographic and semilexicographic choice we confront a complexity-coherence tradeoff. Opting for semilexicographic choice represents a feasible increase in complexity, but also a slightly increased risk of forming incoherent attitudes.

One way to understand this result is to think about *procedural complexity*, the number and complexity of processing steps involved in executing a cognitive process.[7] Feasible increases in procedural complexity often create new opportunities for incoherence by allowing later steps to conflict with earlier ones. In this case, the tiebreaking phase of semilexicographic choice favors a different option than the previous steps favor. We can remove these opportunities for incoherence by refusing to increase procedural complexity. However, increasing procedural complexity may often be a good way to increase expected decision quality by allowing for more extensive and careful analysis of a choice situation. 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

[6]On some interpretations, these binary choices may only reveal weak preferences. Readers who interpret the example in this way are welcome to tweak it using familiar devices, such as mild sweetening, to reveal strict preferences.

[7]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.

cognitive architecture using finite automaton theory, an approach popular in economics and computer science (Oprea 2020; Rubinstein 1986; Salant 2011).

An *automaton* A takes as input lists L from a domain D and chooses an element 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 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*. 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 tells the automaton to stop once an option with utility t is observed, and the output function says to choose that option.[8]

By contrast, expected utility maximization has state complexity $|D|-1$, one less than the cardinality of the option space. One way to implement expected 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.[9] Regrettably, we can prove that no automaton with fewer states implements expected utility maximization (Salant 2011).

In many circumstances, expected 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 expected 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.

[8]That is, $g(s_0, x) = \textsc{stop}$ if $u(x) \geq t$ and $g(s_0, x) = s_0$ otherwise, with $f(s_0, x) = x$.
[9]I.e. $s(i, x)$ returns x if $u(x) > u(x_i)$ and otherwise returns x_i.

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 expected-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 utility-maximizing solution to the choice design problem is *K-phase satisficing* (Salant 2011).

K-phase satisficing is a generalized version of satisficing which affords the agent $K - 1$ opportunities to adjust her utility threshold based on experience. Formally, K-phase satisficing begins with an initial threshold t_0 and a sequence of $K - 1$ *pivotal alternatives* $a_1, \dots, a_{K-1}$ 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 $s_0, \dots, s_{K-1}$ 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 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.[10] 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. When agents cannot afford the state complexity of expected utility maximization, they can still make better expected decisions by shifting from satisficing to K-phase satisficing.

[10]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.

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 expected 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 $t_0, \ldots, t_{K-1}$. 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.[11] The lesson is that computationally tractable forms of aspiration adaptation during satisficing-style decisionmaking are often good ways to improve decision quality. Although aspiration adaptation may represent a desirable increase in complexity, it often induces an accuracy-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.

5 Valence-sensitive inference and informational complexity

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; Rakow and Newell 2010; Wulff et al. 2018). In particular, in many contexts agents respond more coherently when information is presented experientially rather than descriptively (Lejarraga et al. 2016; Schulze and Hertwig 2021; Wulff et al. 2018). This gap

[11] 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.

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 decisionmaking. 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 decisionmaking: the amount and complexity of information used during decisionmaking.[12] I show how potentially desirable ways of increasing informational complexity come at the cost of heightened vulnerability to incoherence, generating an accuracy-coherence tradeoff for agents who must decide whether to increase the informational complexity of their decisionmaking 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.

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.[13] Many theorists will be comfortable taking the incoherence of framing effects on board as a plausible observation about coherence. But 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 $o' \not\succ_S o$.

[12] 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.

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

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.[14] Alternatively, we might with some justification regard the agent as having an irreflexive strict preference for an item over itself, since '80% lean' and '20% fat' meat are the same thing.

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).[15] 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; Krishnamurthy et al. 2001; Payne et al. 2013).[16] 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 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.

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

[15] 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; Lakshminarayanan et al. 2011; Marsh and Kacelnik 2002), and again these effects are controversial (Houston and Wiesner 2020; Kanngiesser and Woike 2016).

[16] In some cases other phenomena such as construal level (Freiling et al. 2014) and subjective scales and experience (Janiszewski et al. 2003) may also play a role.

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? In the next section, I construct a simple model of a choice situation where the price may be worth paying. The model is self-contained, but proofs are left for the appendix.

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 $\textsc{Lean}$ 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 decisionmaking.

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 (Section 3), aspiration adaptation (Section 4) and informational complexity (Section 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 (Section 3) or an increase in the state-complexity of cognitive processes (Section 4). And high levels of complexity allow humans to reap the benefits of symbolic knowledge and understanding, 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 (Geman et al. 1992; 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 Heuristics and incoherence

The *adaptive toolbox* tradition holds that humans have access to a toolbox of fast-and-frugal heuristics such as lexicographic and semilexicographic choice, and that it is often rational for humans to use heuristics (Gigerenzer and Selten 2001; Gigerenzer and Gaissmaier 2011). Against the rationality of heuristics, it is sometimes objected that heuristics are vulnerable to incoherence (Gilovich et al. 2002; Kahneman et al. 1982).[17] But if the discussion in this paper is on the right track, incoherence is not merely a problem for simple heuristics. In many cases, fans of complex cognitive methods must also be prepared to accept a heightened risk of incoherence in exchange for the benefits of cognitive complexity.

This finding suggests that many theorists may have good reason to relax the idea that a vulnerability to incoherence is always to be treated as a decisive objection to the use of specific cognitive processes. Can theorists in the adaptive toolbox tradition develop this thought in order to defend the rationality of heuristic cognition?

6.3 Complications for dual-process accounts

One of the best-known and most controversial (Keren and Schul 2009; Kruglanski and Gigerenzer 2011; Melnikoff and Bargh 2018) paradigms in cognitive science is the *dual process* approach, which divides fast, intuitive and error-prone System 1 thinking from slow, reflective and normative System 2 thinking (Evans et al. 2003; Evans and Stanovich

[17] Sometimes this criticism is walked back a half step. For example, the editors of a recent anthology on heuristics and biases hold: “although … heuristics are distinguished from normative reasoning processes by biased judgments, the heuristics themselves as sensible estimation procedures that are by no measure ‘irrational’.” (Gilovich and Griffin 2002, p. 3). The force of this criticism will depend on what is meant by calling a reasoning process normative and how this differs from calling it rational. See Thorstad (forthcoming b) for discussion.

2013; Sloman 1996). A recent strategy for complicating the distinction between System 1 and System 2 thinking has been to show that many of the dichotomies used to separate System 1 and System 2 thinking are subject to systematic reversals. For example, the 'fast logic' program of Wim De Neys and collaborators contends that System 2 thinking can often be fast, whereas System 1 thinking can often be slow (Bago and De Neys 2017; De Neys forthcoming; Newman et al. 2017).

We might productively take the complexity-coherence tradeoff to problematize another dichotomy associated with the dual process approach. Many theorists hold that System 2 is normative, in the sense of respecting coherence requirements, whereas System 1 is error-prone, in the sense of violating coherence requirements (Evans and Stanovich 2013; Kahneman 2011). But the discussion in this paper suggests that the opposite may often be true. Complex, descriptive and symbolic System 2 thinking can be a source of incoherence which simpler, experience-based System 1 thinking avoids. And this is not a marginal or infrequent occurrence. We began this paper with the insight that there is a strikingly consistent inverse correlation between the complexity and coherence of cognitive systems found in nature. Could this insight be used to put further pressure on traditional ways of separating System 1 and System 2 cognition?

6.4 Approximate coherentism

Traditional theories of rationality require agents to be fully coherent. By contrast, scientific theories of bounded rationality often give only a limited role to coherence (Gigerenzer 2019; Gigerenzer and Sturm 2012). Things are otherwise in philosophy, where a number of Bayesian theorists have defended *approximate coherentism*, the view that bounded agents should strive to be as coherent as possible given their limitations (Staffel 2020; Zynda 1996).

Approximate coherentism has faced pushback from a number of quarters (Arkes et al. 2016; Babic 2021; Daoust forthcoming). One relevant strategy turns on the existence of tradeoffs. Thorstad (forthcoming a) argues that there is an *accuracy-coherence tradeoff* between accuracy and coherence as cognitive goals for bounded agents. Thorstad suggests that bounded agents should sometimes opt for accuracy over coherence. We might strengthen Thorstad's argument by adding a second tradeoff between complexity and coherence as cognitive goals. Could careful reflection on the relationship between complexity and coherence push us away from approximate coherentism, and towards theories of bounded rationality which gives a less central role to coherence?

6.5 Concluding thoughts

I hope that this discussion helps to motivate the idea that there could be a systematic complexity-coherence tradeoff in cognition and to show 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.

Appendix

In the model of Section 5.3, we have:

$$V(x) = q + h - 1.$$
$$\overline{q} = q + \epsilon.$$

Where q, h, ϵ are independent; $q \sim \mathcal{N}(0, 3)$; $\epsilon \sim \mathcal{N}(0, 2)$; and h is Bernoulli with $Pr(h = 0) = Pr(h = -2) = 0.5$.

The quality-only method: Optimal choice

The optimal policy involves choosing bars with quality signal such that:

$$E[q + h - 1|\overline{q}] \geq 0. \tag{1}$$

By linearity of expectations this requires:

$$E[q|\overline{q}] \geq E[1 - h|\overline{q}]. \tag{2}$$

By independence of h, q, ϵ, we have $E[1 - h|\overline{q}] = 1 - E[h] = 2$, giving:

$$E[q|\overline{q}] \geq 2. \tag{3}$$

Since q, ϵ are independent normal distributions with mean 0, we have $E[q|\overline{q}] = E[q|q + \epsilon] = \overline{q}\sigma_q^2/(\sigma_q^2 + \sigma_\epsilon^2) = 9\overline{q}/13$. So optimal choice requires a quality threshold for which $9\overline{q}/13 \geq 2$, or $\overline{q} \geq 26/9$.

The quality-only method: Expected gain

Over all possible candy bars, the expected gain of the quality-only method with a choice threshold of 26/9 is:

$$\int_{26/9}^{\infty} \overline{q}(x) E[q + h - 1|\overline{q} = x] dx. \tag{4}$$

As above,

$$\begin{aligned} E[q + h - 1|\overline{q} = x] &= E[q|\overline{q} = x] - 2 \\ &= 9x/13 - 2. \end{aligned} \tag{5}$$

Recall that for independent normal distributions, $\mathcal{N}(\mu_1, \sigma_1^2) + \mathcal{N}(\mu_2, \sigma_2^2) \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, and hence in particular:

$$\overline{q} \sim q + \epsilon \sim \mathcal{N}(0 + 0, 2 + 3) \sim \mathcal{N}(0, 5). \tag{6}$$

Substituting into (5) and (6) into (4) gives an expected gain of:

$$\frac{1}{5\sqrt{2\pi}} \int_{26/9}^{\infty} e^{-1/2(x/5)^2} [9x/13 - 2] dx. \tag{7}$$

or approximately 0.605.

The label method: Optimal choice

The label method involves observing the random variable $\overline{h}$ which takes values 'lean' and 'fat', with $Pr(h = 0|\overline{h} = \text{lean}) = .75$ and $Pr(h = 0|\overline{h} \neq \text{lean}) = .25$. Here $\overline{h}$ is independent of q, ϵ.

The optimal policy involves choosing candy bars with quality signal such that:

$$E[q + h - 1|\overline{q}, \overline{h}] \geq 0. \tag{8}$$

As before, by linearity of expectations and independence of h, q, ϵ this simplifies to:

$$E[q|\overline{q}] \geq 1 - E[h|\overline{h}]. \tag{9}$$

As before, the left-hand expression is $9\overline{q}/13$, and by specification of $\overline{h}$ our requirement becomes:

$$9\overline{q}/13 \geq \begin{cases} 3/2 & h = \text{lean} \\ 5/2 & h = \text{fat} \end{cases} \tag{10}$$

Simplifying gives:

$$\overline{q} \geq \begin{cases} 13/6 & h = \text{lean} \\ 65/18 & h = \text{fat} \end{cases} \tag{11}$$

The label method: Expected gain

By construction of $h, \overline{h}$, candy bars are equally likely to be labeled 'lean' or 'fat'. This means we can average the gains of the label method across label types to determine the expected gain of the label method.

For bars labeled 'lean' the label method uses a choice threshold of $\overline{q} \geq 13/6$, with expected gain:

$$\int_{13/6}^{\infty} \overline{q}(x) E[q + h - 1|\overline{q} = x, \overline{h} = \text{'lean'}]dx. \tag{12}$$

By linearity of expectations and independence of q, h, ϵ, we have:

$$\begin{aligned} E[q + h - 1|\overline{q} = x, \overline{h} = \text{'lean'}] &= E[q|\overline{q} = x] + E[h|\overline{h} = \text{'lean'}] - 1 \\ &= 9x/13 - 3/2. \end{aligned} \tag{13}$$

And as before, $\overline{q} \sim \mathcal{N}(0, 5)$, so that (12) simplifies to:

$$\frac{1}{5\sqrt{2\pi}} \int_{13/6}^{\infty} e^{-1/2(x/5)^2} [9x/13 - 3/2]dx. \tag{14}$$

which works out to approximately 0.759.

By symmetry, for bars labeled 'fat' the label method uses choice threshold $\overline{q} \geq 65/18$ with expected gain:

$$\frac{1}{5\sqrt{2\pi}} \int_{65/18}^{\infty} e^{-1/2(x/5)^2} [9x/13 - 5/2] dx. \tag{15}$$

which works out to approximately 0.476. Averaging across cases gives an expected gain of 0.618 for the label method.

The deductive method: Optimal choice

The optimal policy under the deductive method involves choosing bars with quality signal such that:

$$E[q + h - 1 | \overline{q}, h] \geq 0]. \tag{16}$$

By independence and linearity, this is the requirement that

$$E[q|\overline{q}] \geq 1 - E[h|h] = 1 - h. \tag{17}$$

As before $E[q|\overline{q}] = 9\overline{q}/13$ so that we need:

$$9\overline{q}/13 \geq 1 - h = \begin{cases} 1 & h = 0 \\ 3 & h = -2. \end{cases} \tag{18}$$

or:

$$\overline{q} \geq \begin{cases} 13/9 & h = 0 \\ 13/3 & h = -2. \end{cases} \tag{19}$$

The deductive method: Expected gain

Since the cases $h = 0$ and $h = -2$ are equally likely, we can assess the deductive method by averaging its performance across cases.

For bars with $h = 0$, the deductive method gives average gain:

$$\int_{13/9}^{\infty} \overline{q}(x) E[q + h - 1 | \overline{q} = x, h = 0] dx. \tag{20}$$

By linearity and independence, we have:

$$\begin{aligned} E[q + h - 1 | \overline{q} = x, h = 0] &= E[q|\overline{q} = x] + E[h - 1 | h = 0] \\ &= 9x/13 - 1. \end{aligned} \tag{21}$$

And since $\overline{q} \sim \mathcal{N}(0, 5)$ the deductive method gives average gain in this case of:

$$\frac{1}{5\sqrt{2\pi}} \int_{13/9}^{\infty} e^{-1/2(x/5)^2} [9x/13 - 1] dx. \tag{22}$$

or approximately 0.938. A symmetrical calculation shows that the deductive method has average gain 0.369 in the case that $h = -2$, for an average gain of 0.654.

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.

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

Babic, Boris. 2021. "Approximate coherentism and luck." *Philosophy of Science* 88:707–25.

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 Rolf Herken (ed.), *The universal Turing machine: A half-century survey*, 227–57. Oxford University Press.

Bermúdez, José. 2020. *Frame it again*. 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.

Crutchfield, James and Young, Karl. 1989. "Inferring statistical complexity." *Physical Review Letters* 65:105.

Daoust, Marc-Kevin. forthcoming. "The comparison problem for approximating epistemic ideals." *Ratio* forthcoming.

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

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.

Evans, Jonathan, Handley, Simon, and Over, David. 2003. "Conditionals and conditional probability." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:321.

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

Fishburn, Peter. 1974. "Lexicographic orders, utilities and decision rules: A survey." *Management Science* 20:1442–71.

Freiling, Traci, Vincent, Leslie, and Henard, David. 2014. "When *not* to accentuate the positive: Re-examining valence effects in attribute framing." *Organizational Behavior and Human Decision Processes* 124:95–109.

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. 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* 1:107–143.

Gigerenzer, Gerd and Gaissmaier, Wolfgang. 2011. "Heuristic decision making." *Annual Review of Psychology* 62:451–82.

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 and Griffin, Dale. 2002. "Heuristics and biases: Then and now." In Thomas Gilovich, Dale Griffin, and Daniel Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 1–18. Cambridge University Press.

Gilovich, Thomas, 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.

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 descriptoin: Exploring two paths to knowledge." *Current Directions in Psychological Science* 27:123–8.

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

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.

Janiszewski, Chris, Silk, Tim, and Cooke, Alan. 2003. "Different scales for different frames: The role of subjective scales and experience in explaining attribute-framing effects." *Journal of Consumer Research* 30:311–25.

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, 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.

Kaplan, Rachel and Kaplan, Stephen. 1989. *The experience of nature: A psychological perspective*. Cambridge University Press.

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.

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

Krishnamurthy, Parthasarathy, Carter, Patrick, and Blair, Edward. 2001. "Attribute framing and goal framing effects in health decisions." *Organizational Behavior and Human Decision Processes* 85:382–99.

Kruglanski, Arie and Gigerenzer, Gerd. 2011. "Intuitive and deliberate judgments are based on common principles." *Psychological Review* 118:97–109.

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, Lambert, James, and Wiesner, Karoline. 2013. "What is a complex system?" *European Journal for Philosophy of Science* 3:33–67.

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

Lakshminarayanan, Venkat, Chen, M. Keith, and Santon, Laurie R. 2011. "The evolution of decision-making under risk: Framing effects in monkey risk preferences." *Journal of Experimental Social Psychology* 47:689–93.

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

Lejarraga, Tomás, Pachur, Thorsten, Frey, Renato, and Hertwig, Ralph. 2016. "Decisions from experience: From monetary to medical gambles." *Journal of Behavioral Decision Making* 29:67–77.

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.

Liu, Peng and Li, Zhizhong. 2012. "Task complexity: A review and conceptualization framework." *International Journal of Industrial Ergonomics* 42:553–68.

Marsh, Banabay and Kacelnik, Alex. 2002. "Framing effects and risky decisions in starlings." *Proceedings of the National Academy of Sciences* 99:3352–5.

McAllister, James. 2003. "Effective complexity as a measure of information content." *Philosophy of Science* 70:302–7.

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.

Newman, Ian, Gibb, Maia, and Thompson, Valerie. 2017. "Rule-based reasoning is fast and belief-based reasoning can be slow: Challenging current explanations of belief-bias and base-rate neglect." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43:1154–70.

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.

Payne, John, Sagara, 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.

Rakow, Tim and Newell, Ben. 2010. "Degrees of uncertainty: An overview and framework for future research on experience-based choice." *Journal of Behavioral Decision Making* 23:1–14.

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 rationality." *American Economic Review* 101:724–48.

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

Schuck-Paim, Cynthia. 2002. "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. 1998. "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.

Sloman, Steven. 1996. "The empirical case for two systems of reasoning." *Psychological Bulletin* 119:3–22.

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.

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

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

—. forthcoming b. "Two paradoxes of bounded rationality." *Philosophers' Imprint* forthcoming.

—. forthcoming c. *Inquiry under bounds*. Oxford University Press.

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

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

Wheeler, Gregory. 2020. "Less is more for Bayesians, too." In Riccardo Viale (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.

Worsnip, Alex. 2021. *Fitting things together: Coherence and the demands of structural rationality*. Oxford University Press.

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.