The complexity-coherence tradeoff in cognition

Abstract

I argue that bounded agents face a systematic complexity-coherence tradeoff in cognition. Increasing the complexity of cognitive processes often induces a heightened risk of incoherence. I illustrate how the complexity-coherence tradeoff can arise as a result of efforts to make cognition more sensitive to complex variation in decision-relevant outcomes. I use a case study of the role of semantic valence in attribute framing to show how this mechanism generates a complexity-coherence tradeoff in practice as well as in theory. I conclude with a discussion of descriptive and normative consequences of the complexity-coherence tradeoff.

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, the developmental psychologist Alison Gopnik wonders: "Why are grown-ups often so stupid about probabilities when even babies and chimps can be so

[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).

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 tradeoff in cognition (Stanovich 2013; Thorstad forthcoming). 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). Next, I argue that the complexity-coherence tradeoff often obtains, and then catalog some of the factors which drive the complexity-coherence tradeoff (Sections 3-9). Finally, I draw out implications of the complexity-coherence tradeoff and sketch directions for future work (Section 10).

2 Clarifying the complexity-coherence tradeoff

What does it mean to speak of a complexity-coherence tradeoff in cognition? We can get a handle on what this might mean 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 three 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.[2] 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.

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

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

of complexity, increasing complexity often comes at the expense of 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. With these clarifications in mind, let us begin with an example designed to illustrate how a complexity-coherence tradeoff could arise.

3 Warmup: Lexicographic and semilexicographic choice

To see how a complexity-coherence tradeoff could arise, it may help to borrow an example from David Thorstad (forthcoming). 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.[3] 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

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 A. It is natural to interpret this as a pattern of intransitive binary preferences.

Thorstad (forthcoming) 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.

[3]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.

I hope that this example goes some way towards illustrating why there might be a complexity-coherence tradeoff in cognition. But I do not want to rest content with a single example. After all, examples can be misleading, and one example does not illustrate a systematic tradeoff. In the rest of this paper, I aim to identify some general cognitive factors which could underly a more systematic tradeoff between complexity and coherence in cognition.

4 Sensitivity to outcome variation

A good place to begin is by developing an insight by Keith Stanovich (2013). Stanovich's insight is that many of the preference axioms enforce a type of context-insensitivity. For example, the Independence of Irrelevant Alternatives requires preferences to be unaffected by the addition of irrelevant options to the agent's choice set.

Stanovich suggests that many ways of making cognition more context-sensitive involve added cognitive complexity. This is particularly plausible when context-sensitivity is purchased by adding new parameters to cognitive processes or by enriching those processes in specifiable ways. All of my examples will have this form.

However, Stanovich argues that making cognition more context-sensitive often carries a heightened risk of incoherence. That is because feasible strategies for making cognition sensitive to relevant types of contextual variation also carry some risk of sensitivity to irrelevant contextual variation of the type forbidden by the preference axioms. When this is the case, decisions to increase context-sensitivity will induce a complexity-coherence tradeoff in cognition.

Making out this argument in full generality would be a harrowing task because there are many things that can be meant by context-sensitivity. In this paper, I focus on one type of context-sensitivity: sensitivity to variation in the features of outcomes, which I will abbreviate as *sensitivity to outcome variation*. I argue that feasible strategies for increasing sensitivity to outcome variation are often subject to a complexity-coherence tradeoff, trading a useful sensitivity to complex variation in outcomes for an increased risk of incoherence.

To see how sensitivity to outcome variation can lead to incoherence, suppose that you take me for a ride. First, you trade me a horse in exchange for my car. Next, you trade me a cow for the horse. Finally, you sell me back my car for the cow, less a hundred-dollar fee. Have I exhibited preferences that are intransitive, and therefore incoherent?

We can sometimes restore coherence by individuating the outcomes of choice more finely.[4] That is, we can find some feature F which varies between the outcomes of my choices, but which was not described in my presentation of the problem. Then we can redescribe my preferences by building in the presence or absence of F. For example, suppose that F is the presence of a fuel shortage, and suppose that F obtained until the day I bought my car back. Then my preferences over outcomes take the following form:

$$(\text{car}, F), \prec (\text{horse}, F), \prec (\text{cow}, F) \prec (\text{car} - \$100, \neg F).$$

These preferences are not incoherent.

Now it is nearly always possible to find some F which differs between outcomes and to claim that my preferences are perfectly coherent because they are sensitive to variation in F. For example, suppose that the first two trades take place on a Monday and let F be the property of being Monday. And suppose this is the only relevant difference between outcomes. It is not that I would be able to make better use of a horse on Mondays, or that I would experience more positive hedonic feelings on a Monday-morning ride. I simply prefer the outcome (horse, Monday) to (car, Monday), but prefer (car, Tuesday) to (horse, Tuesday). Does this mean that my preferences are actually coherent, so long as they genuinely vary with the day of the week?

Many authors have stood their ground and held that such preferences are incoherent, in

[4]In this discussion I pass freely between talk of the objects of choice (options) and the outcomes of choice. I do this because my examples involve decisionmaking under certainty, so that options are constant functions from states to a fixed outcome. To the best of my knowledge, all of the arguments in this paper can be rephrased to cover choice between chancy prospects.

the sense of not living up to the normative requirements of the preference axioms (Broome 1991; Dreier 1996; Pettit 1991). It is not coherent to have a preference between outcomes unless there is a relevant difference between them which could ground a preference. For example, John Broome (1991) suggests that it is incoherent to have a preference between outcomes unless they differ in some *justifier* which makes it rational to have a preference between them. My purpose here is not to endorse one relevance criterion, such as Broome's individuation by justifiers, over others. It is rather to suggest that many authors would endorse a coherence requirement of the following form:

> **Insensitivity to Irrelevant Outcome Variation (IIOV):** For all agents S and outcomes o, o', if S prefers o to o' then there must be some relevant difference in the features of o and o'.

Most nontrivial ways of filling out the relevance requirement in IIOV will suffice to drive the argument of this paper.

5 Asymmetric responsibility for outcome variation

A notable feature of IIOV is that it enforces an asymmetry of responsibility for sensitivity to the varying features of outcomes. Agents can become incoherent by taking account of irrelevant variation in features of outcomes. For example, it is incoherent to prefer (horse, Monday) to (car, Monday) while preferring (car, Tuesday) to (horse, Tuesday). But agents cannot become incoherent by failing to take account of important features of outcomes. It is not incoherent to prefer the receipt of any horse to the receipt of any car, without regard to the features of the horse or the car.

This asymmetric responsibility for outcome variation will generate a complexity-coherence tradeoff if we grant two further assumptions. I give only preliminary motivations for each assumption here, since the rest of this paper will exhibit a class of cases in which both assumptions can be plausibly said to hold.

The first assumption is that agents often have reason to make their cognition more sensitive to outcome variation. Suggestive evidence for this assumption would be provided by the *evolutionary complexity thesis* in evolutionary psychology, on which the function of cognitive complexity is to enable agents to cope with environmental complexity (Godfrey-Smith 2002). Making our cognition more complex can be an efficient way to flexibly adapt behavior to the demands of changing environments (Spurrett 2021).[5] And in particular, making ourselves sensitive to additional forms of outcome variation allows us to respond differently to environments which present us with relevantly different outcomes to choose between.

The first assumption means that the complexity-coherence tradeoff cannot be avoided by making ourselves maximally insensitive to outcome variation, since we sometimes have strong reason to become more sensitive to outcome variation. That fact will produce a complexity-coherence tradeoff if we grant a second assumption: that feasible strategies for increasing sensitivity to outcome variation often carry a heightened risk of incoherence. IIOV ensures that agents cannot become less coherent by making themselves insensitive to outcome variation. But agents risk becoming less coherent by making themselves more sensitive to outcome variation. Even if, in principle, some highly complex strategies could incorporate additional outcome variation without incoherence, strategies of feasible complexity often increase an agent's sensitivity to relevant outcome variation at a small risk of sensitivity to irrelevant outcome variation. It is not always feasible to ensure that only relevant variation will be incorporated.

If that is right, then in choosing whether to increase their sensitivity to outcome variation, agents confront a complexity-coherence tradeoff. Increased complexity, in the form of increased sensitivity to outcome variation, may carry worthwhile benefits, but nonetheless heightens the risk of forming incoherent attitudes. That is the theory. But could this happen in practice?

[5]Theories of ecological rationality stress that sometimes simple methods are the best way to respond to rapidly changing environments (Todd and Gigerenzer 2012; Berg and Watanabe 2020). This is fully compatible with my view. To say that complexity is sometimes beneficial is not to make the stronger claim that more complexity is always better. I return to this discussion in Section 10.

6 Complexity, coherence and the description-experience gap

To see how increased sensitivity to outcome variation could lead in practice to incoherence, consider the description-experience gap (Hertwig et al. 2004). 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 much more coherently when information is presented experientially rather than descriptively (Lejarraga et al. 2016; Schulze and Hertwig 2021; Wulff et al. 2018). This gap in responding to described versus experienced information is known as the *description-experience gap*.

What does the description-experience gap have to do with the complexity-coherence tradeoff? To see how it might be relevant, 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, and because learning through description raises the risk of incoherent responding, humans are often more incoherent than nonhuman animals, who never learn through description. Consistent with this explanation, when humans are tested in the same experimental paradigms as nonhuman animals, being provided with experiential rather than descriptive information, humans respond just as coherently as nonhumans do (Hertwig et al. 2018; Schulze and Hertwig 2021).

This discussion suggests that the description-experience gap may be a good way to illustrate how cognitive complexity leads to incoherence. But before we single out complexity as the culprit, we need to ask what causes the description-experience gap. And here two qualifications are in order. First, the description-experience gap may not arise for all agents or across all paradigms (Camilleri and Newell 2013; Glöckner et al. 2012). And second, there are many causes of the description-experience gap, some of which may not have much to do with the complexity-coherence tradeoff (Hertwig and Erev 2009; Wulff et al. 2018).

Nevertheless, our discussion of human and nonhuman animals suggests one natural way in which a description-experience gap can arise. The suggestion is that descriptive information facilitates or enables more complex forms of cognitive responding, which in turn involve a heightened risk of incoherence.[6] In this paper, I focus on a particular form of complex responding identified in Sections 4-5: increased sensitivity to outcome variation. How could descriptive information facilitate sensitivity to outcome variation?

One possibility is that descriptive information enables the presentation of decision cues which, although predictive of outcome quality, are not themselves valuable features of outcomes. In Sections 7-8, I discuss one such cue: the semantic valence of descriptions used to present outcomes. Because cues such as semantic valence are correlated with relevant variation in outcomes, agents may often have good reason to use them during decisionmaking. However, because cues such as semantic valence are not themselves features of outcomes, they are sometimes sensitive to spurious outcome variation. That is, they may vary without any genuine change in relevant features of outcomes. This creates the problem of *cue-feature mismatch*: feasible strategies for incorporating many descriptive cues run a risk of producing changes in preference without any change in relevant features of outcomes. In such cases, we saw in Section 5 that agents confront a complexity-coherence tradeoff: feasible strategies for incorporating cues such as semantic valence present a sometimes-desirable increase in complexity, but an increased risk of incoherence.

[6]In this vein, Glöckner et al. (2012) and Hertwig and Erev (2009) among others suggest that processing differences may partially underly the description-experience gap. And Lejarraga (2010) and Weiss-Cohen et al. (2018) show that agents sometimes prefer experience to description in highly complex problems, suggesting that descriptive responding is often more complex.

How does the problem of cue-feature mismatch arise in practice? To illustrate the point, Sections 7-8 examine the role of semantic valence in framing effects.

7 Attribute framing and the valence-consistent shift

Framing effects occur when agents take different attitudes towards equivalent presentations of the same 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. Many framing effects involve incoherence of preference. For example, both of the cases just described involve irreflexive strict preference for an outcome over a different presentation of the same outcome.

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). Although some framing effects have been alleged in response to experiential information (Fu et al. 2018; Gonzalez and Mehlhorn 2016), these effects 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).[7]

One of the most-studied categories of framing effects is *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

[7]These effects are often controversial (Houston and Wiesner 2020; Kanngiesser and Woike 2016).

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).[8] 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, and hence relying on semantic valence may increase agents' sensitivity to relevant outcome variation. 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 responding to spurious outcome variation, because semantic valence is subject to cue-feature mismatch. 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

[8]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.

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.

8 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 $Pr(h = 0|\text{LEAN}) = .75$ and $Pr(h = 0|\neg\text{LEAN}) = .25$.

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 startlingly 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.

9 Taking stock

In this paper, I have argued that there is a complexity-coherence tradeoff in cognition. In expectation and within a range of feasible strategies, opting for more complex cognitive processes often decreases the expected coherence of an agent's resulting attitudes. I illustrated the complexity-coherence tradeoff by looking at a special type of context-sensitivity, sensitivity to outcome variation. Requirements of coherence enforce an asymmetry of normative responsibility for outcome variation. They require agents to avoid sensitivity to irrelevant variation in outcome features, but do not penalize agents for failing to be sensitive to relevant variation in outcome features.

This asymmetry of responsibility creates a problem, because agents often have good reason to rely on decision cues that are not themselves valuable features of outcomes, or whose relevance to decisionmaking is not exhausted by their value as outcome features. That creates the possibility of cue-feature mismatch, in which decision cues vary without relevant variation in outcomes. In that case, a complexity-coherence tradeoff arises. A feasible way for agents to make their cognition more complex is to rely on decision cues which are vulnerable to cue-feature mismatch. But because of this vulnerability, relying on additional cues makes agents more likely to form incoherent attitudes.

I illustrated the problem of cue-feature mismatch through a case study of attribute framing. Attribute framing occurs when the semantic valence of an attribute is varied across otherwise equivalent descriptions of an object. Because semantic valence is a predictive decision cue, agents often rely on semantic valence to guide decisionmaking. However, since semantic valence is not itself a valuable feature of outcomes, it is subject to cue-feature mismatch. This mismatch is laid bare in attribute framing manipulations, which induce a change of preference by manipulating semantic valence without relevant outcome variation.

If this is right, then it lends some credibility to the idea that there could be a more general complexity-coherence tradeoff in cognition and illustrates some of the factors driving the complexity-coherence tradeoff. In the next section, I discuss how agents might confront the complexity-coherence tradeoff and use this discussion to highlight open questions for future research.

10 Future directions

So far, we have seen that complexity and coherence conflict as cognitive desiderata. Sometimes opting for more complex cognitive processes reduces the expected coherence of the attitudes that will result. How might agents confront the complexity-coherence tradeoff? This choice is deeply fraught. Thinking through the ways in which the choice between complexity and coherence is fraught will reveal three promising directions for future research.

Against complexity, it is traditionally urged that complex processes are often slow and cognitively costly. Moreover, the traditional case continues, 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). For example, if we are willing to pursue the thought that cognitive processes based on descriptive information are often more complex than cognitive processes based on experiential information (Glöckner et al. 2012), then the description-experience gap gives us some reason to question the doctrine that more complex processes are always better. After all, research on the description-experience gap shows that in several paradigms, experience-based cognition outperforms descriptive cognition, not merely in coherence, but also on criteria such as judgmental accuracy and decision quality (Lejarraga 2010; Lejarraga and Hertwig 2021).

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. Coherence may be intrinsically valuable, and it is certainly instrumentally valuable. Insofar as this is so, agents have yet another reason to resist using complex processes.

However, in many situations there are definite advantages to cognitive complexity. Sometimes, making processes more complex is a good way to increase decision quality at a feasible cost, as in the turn from lexicographic to semilexicographic choice (Section 3) or the use of semantic valence as a decision cue (Sections 7-8). 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.

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.

10.1 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).[9] But if the discussion in this paper is on the right track, incoherence is not merely a problem for the adaptive toolbox tradition. 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?

10.2 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 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

[9]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.

can often be fast, whereas System 1 thinking can often be slow (Bago and De Neys 2017; 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?

10.3 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) 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?

10.4 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 8, 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:

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

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.

Berg, Nathan and Watanabe, Yuki. 2020. "Conservation of behavioral diversity: On nudging, paternalism-induced monoculture, and the social value of heterogeneous beliefs and behavior." *Mind and Society* 19:103–20.

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

Broome, John. 1991. *Weighing goods: Equality, uncertainty and time*. Blackwell.

Camilleri, Adrian and Newell, Ben. 2013. "Mind the gap? Description, experience, and the continuum of uncertainty in risky choice." *Progress in Brain Research* 202:55–71.

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.

Dreier, James. 1996. "Rational preference: Decision theory as a theory of practical rationality." *Theory and Decision* 40:249–76.

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.

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.

Glöckner, Andreas, Fiedler, Susanna, Hochman, Guy, Ayal, Shahar, and Hilbig, Benjamin. 2012. "Processing differences between descriptions and experience: A comparative analysis using eye-tracking and physiological measures." *Frontiers in Psychology* 3:1–15.

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

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, Barron, Greg, Weber, Elke, and Erev, Ido. 2004. "Decisions from experience and the effect of rare events in risky choice." *Psychological Science* 15:534–9.

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.

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.

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.

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. 2010. "When experience is better than description: Time delays and complexity." *Journal of Behavioral Decision Making* 23:100–16.

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.

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

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

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.

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.

Pettit, Phillip. 1991. "Decision theory and folk psychology." In Michael Bacharach and Susan Hurley (eds.), *Foundations of decision theory: Issues and advances*. Blackwell.

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.

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.

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.

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.

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

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

Todd, Peter and Gigerenzer, Gerd. 2012. *Ecological rationality: Intelligence in the world*. 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.

Weiss-Cohen, Leonardo, Konstantinidis, Emmanouil, Speekenbrink, Maarten, and Harvey, Nigel. 2018. "Task complexity moderates the influence of descriptions in decisions from experience." *Cognition* 170:209–27.

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.

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.