A Rational Reinterpretation of Dual-Process Theories

Smitha Milli[a, *], Falk Lieder[b], Thomas L. Griffiths[c]

smilli@berkeley.edu, falk.lieder@tuebingen.mpg.de,

tomg@princeton.edu

[a] Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA, USA 94704

[b] Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4 72076, Tübingen, Germany

[c] Departments of Psychology and Computer Science, Princeton University, Princeton, NJ, 08544

Author Note

[*]Corresponding author. A preliminary version of Simulations 1 and 2 was presented at the Thirty-First AAAI Conference on Artificial Intelligence and appeared in the proceedings of that conference (Milli, Lieder, & Griffiths, 2017). The work presented in this article was supported by grant number ONR MURI N00014-13-1-0341 and a grant from the Future of Life Institute.

Abstract

Highly influential “dual-process" accounts of human cognition postulate the coexistence of a slow accurate system with a fast error-prone system. But why would there be just two systems rather than, say, one or 93? Here, we argue that a dual-process architecture might be neither arbitrary nor irrational, but might instead reflect a rational tradeoff between the cognitive flexibility afforded by multiple systems and the time and effort required to choose between them. We investigate what the optimal set and number of cognitive systems would be depending on the structure of the environment. We find that the optimal number of systems depends on the variability of the environment and the difficulty of deciding when which system should be used. Furthermore, when having two systems is optimal, then the first system is fast but error-prone and the second system is slow but accurate. Our findings thereby provide a rational reinterpretation of dual-process theories.

Keywords: bounded rationality; dual-process theories; meta-decision making; bounded optimality; metareasoning; resource-rationality

A Rational Reinterpretation of Dual-Process Theories

Starting in the 1960s, a number of findings began to suggest that people's judgments and decisions systematically deviate from the predictions of logic, probability theory, and expected utility (Wason, 1968; Tversky & Kahneman, 1974; Kahneman & Tversky, 1979; Gilovich, Griffin, & Kahneman, 2002). These deviations are often referred to as *cognitive biases* and have fueled the heated debate about human rationality (Stanovich, 2009; Gigerenzer, 1991; Kahneman & Tversky, 1996). It is commonly assumed that cognitive biases result from people's use of rather arbitrary *heuristics* (Tversky & Kahneman, 1974; Gilovich et al., 2002), thus leading some to conclude that people are fundamentally irrational (Sutherland, 2013; Marcus, 2009; Ariely, 2009). However, others have argued that many apparent errors in human judgment can be understood as rational solutions to a different construal of the problem participants were presumably trying to solve (Oaksford & Chater, 1994, 2007; Hahn & Oaksford, 2007; Hahn & Warren, 2009; Tenenbaum & Griffiths, 2001; Griffiths & Tenenbaum, 2001; Austerweil & Griffiths, 2011; Parpart, Jones, & Love, 2017).

These rational explanations build on the methodology of *rational analysis* (Anderson, 1990; Chater & Oaksford, 1999), which aims to explain the function of cognitive processes by assuming the human mind is well-adapted to the structure of the environment and the problems people are trying to solve. In other words, rational analysis assumes that the human mind implements a (near) rational solution with respect to the underlying computational problem the mind is trying to solve. A more recent line of work on *resource-rational analysis* extends this idea and assumes that the human mind is well-adapted to problems after taking into account the constraint of limited time or cognitive resources (Lieder & Griffiths, in revision; Griffiths, Lieder, & Goodman, 2015). In other words, resource-rational analysis assumes that the human mind rationally trades-off the benefit of accurate solutions against the limited resources available. Under this framework, when time or cognitive resources are abundant, then it is rational to perform more computation, and when time or cognitive resources are limited, then it is rational to do less computation. In this way, many supposedly ad-hoc heuristics have been reinterpreted as being rational solutions when resources are limited (Lieder, Griffiths, & Hsu, 2018; Lieder, Griffiths, Huys, & Goodman, 2018a, 2018b; Howes, Warren, Farmer, El-Deredy, & Lewis, 2016; Khaw, Li, & Woodford, 2017; Sims, 2003; Tsetsos et al., 2016; Bhui & Gershman, 2017). Furthermore, people appear to *adaptively* choose between their fast heuristics and their slower and more deliberate strategies based on the amount of resources available (Lieder & Griffiths, 2017)

However, an issue still remains unresolved in the push for the resource-rational reinterpretation of these heuristics. Since the exact amount of computation to do for a problem depends on the particular time and cognitive resources available, a larger repertoire of reasoning systems should enable the mind to more flexibly adapt to different situations (Payne, Bettman, & Johnson, 1993; Gigerenzer & Selten, 2002). In fact, achieving the highest possible degree of adaptive flexibility would require choosing from an infinite set of diverse cognitive systems. However, this is not consistent with behavioral and neuroscientific evidence for a small number of qualitatively different decision systems (van der Meer, Kurth-Nelson, & Redish, 2012; Dolan & Dayan, 2013) and similar evidence in the domain of reasoning (Evans, 2003, 2008; Evans & Stanovich, 2013).

One reason for a smaller number of systems could be that as the number of systems increases it becomes increasingly more time-consuming to select between them (Lieder & Griffiths, 2017). This suggests that the number and nature of the mind's cognitive systems might be shaped by the competing demands for the ability to flexibly adapt one's reasoning to the varying demands of a wide range of different situations and the necessity to do so quickly and efficiently. In our work, we theoretically formalize this explanation, allowing us to derive not only what the optimal system is given a particular amount of resources, but what the optimal *set* of systems is for a human to select between across problems.

Such an explanation may provide a rational reinterpretation of *dual-process theories*, the theory that the mind is composed of two distinct types of cognitive systems: one that is deliberate, slow, and accurate, and a second one that is fast, intuitive, and fallible (Evans, 2008; Kahneman & Frederick, 2002, 2005). Similar dual-process theories have independently emerged in research on decision-making (Dolan & Dayan, 2013) and cognitive control (Diamond, 2013). While recent work in these areas has addressed the question of how the mind arbitrates between the two systems (Daw, Niv, & Dayan, 2005; Keramati, Dezfouli, & Piray, 2011; Lieder & Griffiths, 2017; Shenhav, Botvinick, & Cohen, 2013; Boureau, Sokol-Hessner, & Daw, 2015), it remains normatively unclear why the mind would be equipped with these two types of cognitive system, rather than another set of systems.

The existence of the accurate and deliberate system, commonly referred to as *System 2* following Kahneman and Frederick (2002), is easily justified by the benefits of rational decision-making. By contrast, the fast and fallible system (*System 1*) has been interpreted as a kluge (Marcus, 2009) and its mechanisms are widely considered to be irrational (Sutherland, 2013; Ariely, 2009; Tversky & Kahneman, 1974; Gilovich et al., 2002). This raises the question why this system exists at all. Recent theoretical work provided a normative justification for some of the heuristics of System 1 by showing that they are qualitatively consistent with the rational use of limited cognitive resources (Griffiths et al., 2015; Lieder, Griffiths, & Hsu, 2018; Lieder, Griffiths, Huys, & Goodman, 2018a, 2018b) – especially when the stakes are low and time is scarce and precious. Thus, System 1 and System 2 appear to be rational for different kinds of situations. For instance, you might want to rely on System 1 when you are about to get hit by a car and have to make a split-second decision about how to move. But, you might want to employ System 2 when deciding whether or not to quit your job.

Here, we formally investigate what set of systems would enable people to make the best possible use of their finite time and cognitive resources. We derive the optimal tradeoff between the cognitive flexibility afforded by mutliple systems and the cost of choosing between them. To do so, we draw inspiration from the artificial intelligence literature on designing intelligent agents that make optimal use of their limited-performance hardware by building upon the mathematical frameworks of *bounded optimality* (Russell & Subramanian, 1995) and *rational metareasoning* (Russell & Wefald, 1991b; Hay, Russell, Tolpin, & Shimony, 2012). We apply this approach to four different domains where the dual systems framework has been applied to explain human decision-making: binary choice, planning, strategic interaction, and multi-alternative, multi-attribute risky choice. We investigate how the optimal cognitive architecture for each domain depends on the variability of the environment and the cost of choosing between multiple cognitive systems, which we call *metareasoning cost*.

This approach allows us to extend the application of resource-rational analysis from a particular system of reasoning to sets of cognitive systems, and our findings provide a normative justification for dual-process theories of cognition. Concretely, we find that across all four domains the optimal number of systems increases with the variability of the environment but decreases with the costliness determining when which of these systems should be in control. In addition, when it is optimal to have two systems, then the difference in their speed-accuracy tradeoffs increases with the variability of the environment. In variable environments, this results in one system that is accurate but costly to use and another system that is fast but error-prone. These predictions mirror the assertions of dual-process accounts of cognition (Evans, 2008; Kahneman, 2011). Our findings cast new light on the debate about human rationality by suggesting that the apparently conflicting views of dual-process theories and rational accounts of cognition might be compatible after all.

The remainder of this paper is structured as follows: We start by summarizing previous work in psychology and artificial intelligence that our article builds on. We then describe our mathematical methods for deriving optimal sets of cognitive systems. The subsequent four sections apply this methodology to the domains of binary choice, planning, strategic interaction in games, and multi-alternative risky choice. We conclude with the implications of our findings for the debate about human rationality and directions for future work.

Background

Before delving into the details of our analysis, we first discuss how our approach applies to the various dual-process theories in psychology, and how we build on the ideas of bounded optimality and rational metareasoning developed in artificial intelligence research.

Dual-process theories

The idea that human minds are composed of multiple interacting cognitive systems first came to prominence in the literature on reasoning (Evans, 2008; Stanovich, 2011). While people are capable of reasoning in ways that are consistent with the prescriptions of logic, they often do not. Dual-process theories suggested that this is because people employ two types of cognitive strategies: fast but fallible heuristics that are triggered automatically and deliberate strategies that are slow but accurate.

Different dual-process theories vary in what they mean by two cognitive systems. For example, Evans and Stanovich (2013) distinguish between dual processes, in which each process can be made up of multiple cognitive systems, and dual systems, which corresponds to the literal meaning of two cognitive systems. Because our work abstracts these cognitive systems based on their speed-accuracy tradeoff our analysis applies both at the level of systems or processes as long as the systems or processes accomplish speed-accuracy tradeoffs. Thus, our theory still applies to both dual "processes" and dual "systems".

There is also debate over how the two systems would interact. Some theories postulate the existence of a higher-level controller that chooses between the two systems (Norman & Shallice, 1986; Shenhav et al., 2013), some that the two systems run in parallel, and others that the slower system interrupts the faster one (Evans & Stanovich, 2013). The analysis we present simply assumes that there is greater metareasoning cost incurred for each additional system. This is clearest to see when a higher-level controller needs to make the decision of which system to employ. Alternatively, if multiple cognitive systems operated in parallel, the cost of arbitrating between these systems would also increase with the number of systems – just like the metareasoning cost. So, we believe our analysis would also apply under this alternative assumption.

Since their development in the reasoning literature, dual-process theories have been applied to explain a wide range of mental phenonomena, including judgment and decision-making, where it has been popularized by the distinction between System 1 and System 2 (Kahneman &

Frederick, 2002, 2005; Kahneman, 2011), and moral reasoning where the distinction is made between a fast deontological system and a slow utilitarian system (Greene, 2015). In parallel with this literature in cognitive psychology, research on human reinforcement learning has led to similar conclusions. Behavioral and neural data suggest that the human brain is equipped with two distinct decision systems: a fast, reflexive, system based on habits and a slow, deliberate system based on goals (Dolan & Dayan, 2013). The mechanisms employed by these systems have been mapped onto model-based versus model-free reinforcement learning algorithms. A model-free versus model-based distinction has also been suggested to account for the nature of the two systems posited to underlie moral reasoning (Cushman, 2013; Crockett, 2013).

The empirical support for the idea that the human mind is composed of two types of cognitive systems raises the question of why such a composition would evolve from natural selection. Given that people outperform AI systems in most complex real-world tasks despite their very limited cognitive resources (Gershman, Horvitz, & Tenenbaum, 2015), we ask whether being equipped with a fast but fallible and a slow but accurate cognitive system can be understood as a rational adaption to the challenge of solving complex problems with limited cognitive resources (Griffiths et al., 2015).

Bounded Optimality and Resource-Rational Analysis

Recent work has illustrated that promising process models of human cognition can be derived from the assumption that the human mind makes optimal use of cognitive resources that are available to it (Griffiths et al., 2015; Lewis, Howes, & Singh, 2014). This idea can be formalized by drawing on the theory of *bounded optimality* which was developed as a foundation for designing optimal intelligent agents. In contrast to expected utility theory (Von Neumann & Morgenstern, 1944), bounded optimality takes into account the constraints imposed by performance-limited hardware and the requirement that the agent has to interact its environment in real time (Russell & Subramanian, 1995). The basic idea is to mathematically derive a program that would enable the agent to interact with its environment as well as or better than any other program that its computational architecture could execute. Critically, the agent's limited computational resources and the requirement to interact with a potentially very complex, fast-paced, dynamic environment in real-time entail that the agent's strategies for reasoning and decision-making have to be extremely efficient. This rules out naive implementations of Bayes rule and expected utility maximization as those would take so long to compute that the agent would suffer a decision paralysis so bad that it might die before taking even a single action.

The fact that people are subject to the same constraints makes bounded optimality a promising normative framework for modeling human cognition (Griffiths et al., 2015). *Resource-rational analysis* applies the principle of bounded optimality to derive optimal cognitive strategies from assumptions about the problem to be solved and the cognitive architecture available to solve it (Griffiths et al., 2015). Recent work illustrates that this approach can be used to discover the discover and make sense of people's heuristics for judgment (Lieder, Griffiths, Huys, & Goodman, 2018a) and decision-making (Lieder, Griffiths, Huys, & Goodman, 2018a; Lieder, Griffiths, & Hsu, 2018), as well as memory and cognitive control (Howes et al., 2016). The resulting models have shed new light on the debate about human rationality (Lieder, Griffiths, Huys, & Goodman, 2018a, 2018b; Lieder, Krueger, & Griffiths, 2017; Lieder, Griffiths, Huys, & Goodman, 2018b; Lieder, Griffiths, & Hsu, 2018; Griffiths et al., 2015). While this approach has so far focused on one individual strategy at a time, the research presented here extends it to deriving optimal cognitive architectures comprising multiple systems or strategies for a wider range of problems. To do so, we use the theory of rational metareasoning as a foundation for modeling how each potential cognitive architecture would decide when to rely on which system or strategy.

Rational metareasoning as a framework for modeling the adaptive control of cognition

Previous research suggests that people flexibly adapt how they decide to the requirements of the situation (Payne, Bettman, & Johnson, 1988). Recent theoretical work has shown that this adaptive flexibility can be understood within the *rational metareasoning* framework developed in artificial intelligence (Lieder & Griffiths, 2017). Rational metareasoning (Russell & Wefald, 1991b; Hay et al., 2012) formalizes the problem of selecting computations so as to make optimal use of finite time and limited-performance hardware. The adaptive control of computation afforded by rational metareasoning is critical for intelligent systems to be able to solve complex and potentially time-critical problems on performance-limited hardware (Horvitz, Cooper, & Heckerman, 1989; Russell & Wefald, 1991b). For instance, it is necessary for a patient-monitoring system used in emergency medicine to metareason in order to decide when to terminate diagnostic reasoning and recommend treatment. (Horvitz & Rutledge, 1991). This example illustrates that rational metareasoning may be necessary for agents to achieve bounded-optimality in environments that pose a wide range of problems that require very different computational strategies. However, to be useful for achieving bounded-optimality, metareasoning has to be done very efficiently.

In principle, rational metareasoning could be used to derive the optimal amount of time and mental effort that a person should invest into making a decision (Shenhav et al., 2017). Unfortunately, selecting computations optimally is a computation-intensive problem itself because the value of each computation depends on the potentially long sequence of computations that can be performed afterwards. Consequently, in most cases, solving the metareasoning problem *optimally* would defeat the purpose of trying to save time and effort (Lin, Kolobov, Kamar, & Horvitz, 2015; Hay et al., 2012; Russell & Wefald, 1991a). Instead, to make optimal use of their finite computational resources bounded-optimal agents (Russell & Subramanian, 1995) must optimally distribute their resources between metareasoning and reasoning about the world. Thus, studying bounded-optimal metareasoning might be a way to understand how people manage to allocate their finite computational resources near-optimally with very little effort (Gershman et al., 2015; Keramati et al., 2011).

Recent work has shown that approximate metareasoning over a discrete set of cognitive strategies can save more time and effort than it takes and thereby improve overall performance (Lieder et al., 2014). This approximation can drastically reduce the computational complexity of metareasoning while achieving human-level performance (Lieder et al., 2014; Lieder & Griffiths, 2017). Thus, rather than metareasoning over all possible sequences of mental operations to determine the exact amount of time to think, humans may simply metareason over a finite set of cognitive systems that have different speed and accuracy tradeoffs. This suggests a cognitive architecture comprising multiple systems for reasoning and decision making and a executive control system that arbitrates between them – which is entirely consistent with extant theories of cognitive control and mental effort (Norman & Shallice, 1986; Shenhav et al., 2017, 2013). Dual-process theories can be seen as a special case of this cognitive architecture where the number of decision systems is two.

According to this perspective, the executive control system selects between a limited number of cognitive systems by predicting how well each of them would perform in terms of decision quality and effort and then selects the systems with the best predicted performance (Lieder & Griffiths, 2017). Assuming that each of these predictions takes a certain amount of mental effort, this entails that the cost of deciding which cognitive system to rely on in a given situation increases with the number of systems. At the same time, increasing the number of systems also increases the agent's cognitive flexibility thereby enabling it to achieve a higher level of performance across a wider range of environments. Conversely, reducing the space of computational mechanisms the agent can choose from entails that there may be problems for which the optimal computational mechanisms will be no longer available. This dilemma necessitates a tradeoff that sacrifices some flexibility to increase the speed at which cognitive mechanisms can be selected. This raises the question of how many and which computational mechanisms a bounded-optimal metareasoning agent should be equipped with, which we proceed to explore in the following sections.

Deriving Bounded-Optimal Cognitive Systems

We now describe our general approach for extending resource-rational analysis to the level of cognitive architectures. The first step is to model the environment. For the purpose of our analysis, we characterize each environment by the set of decision problems $\mathcal{D}$ that it poses to people and a probability distribution P over $\mathcal{D}$ that represents how frequently the agent will encounter each of them. The set of decision problems $\mathcal{D}$ could be quite varied, for example, it could include deciding which job to pick and deciding what to eat for lunch. In this case P would encode the fact that deciding what to eat for lunch is a more common type of decision problem than deciding which job to pick. Associated with each decision problem d is a utility function $U_d(a)$ that represents the utility gained by the agent for taking action a in decision problem d.

Having characterized the environment in terms of decision problems, we now model how people might solve them. We assume that there is a set of reasoning and decision-making systems $\mathcal{T}$ that the agent could potentially be equipped with. The question we seek to investigate is what subset $\mathcal{M} \subseteq \mathcal{T}$ is optimal for the agent to actually be equipped with. The optimal set of systems $\mathcal{M}$ is dependent on three costs: (1) the *action cost*: the cost of taking the chosen action, (2) the *reasoning cost*: the cost of using a system from $\mathcal{M}$ to reason about which action to take, (3) the *metareasoning cost*: the cost of deciding which system to use to decide which action to take. For simplicity, we will describe each of the costs in terms of time delays, although they also entail additional costs, including metabolic costs.

As an example, consider the scenario of deciding what to order for lunch at a restaurant. The diner has a fixed amount of time she can spend at lunch until she needs to get back to work, so time is a finite resource. The *action cost* is the time required to eat the meal. A person might have multiple systems for deciding which items to choose. For example, one system may rely on habit and order the same dish as last time. Another system may perform more logical computation to analyze the nutritional value of each item or what the most economical choice is. Each system has an associated reasoning cost, the time it takes for that system to decide which item to order.

It is clear that the diner has to balance the amount of time spent thinking about what meal to pick (reasoning cost) with the amount of time it will take to actually eat the meal (action cost), so that she is able to finish her meal in the time she has available. If the diner is extremely time-constrained, perhaps because of an urgent meeting she needs to get back to, then she may simply heuristically plop items onto her plate. But, if the diner has more time, then she may think more about what items to choose.

In addition to the cost of reasoning and the cost of acting, having multiple decision systems also incurs the cost of *metareasoning*, that is reasoning about how to reason about what to do. In other words, the metareasoning cost is how much time it takes the diner to decide how much to think about whether to rely on her habits, an analysis of nutritional value, or any of the the other decision mechanisms she may have at her disposal. If the diner only has one system of thinking, then the metareasoning cost is zero. But as the number of systems increases, the metareasoning cost of deciding which system should be in control increases. This raises the question of what is the optimal ensemble of cognitive systems, how many systems does it include, and what are they? We can derive the answer to these questions by computing minimizing the expected sum of action cost, reasoning cost, and metareasoning cost over the set of all possible ensembles of cognitive systems.

In summary, our approach for deriving a bounded-optimal cognitive architecture proceeds as follows:

1. **Model the environment.** Define the set of decision problems $\mathcal{D}$, the distribution over them P, and the utility for each problem $U_d(a)$.

2. **Model the agent.** Define the set of possible cognitive systems $\mathcal{T}$ the agent could have.

3. **Specify the optimal mind design problem.** Define the metric that the bounded agent's behavior optimizes, i.e., a trade-off between the utility it gains and the costs that it incurs; the action cost, reasoning cost, and metareasoning cost.

4. **Solve the optimal mind design problem.** Solve (3) to find the optimal set of systems $\mathcal{M} \subseteq \mathcal{T}$ for the agent to be equipped with.

Once we have done this, we can begin to probe how different parts of the simulation affect the final result in step (4). For example, we expect that the optimal cognitive architecture for a

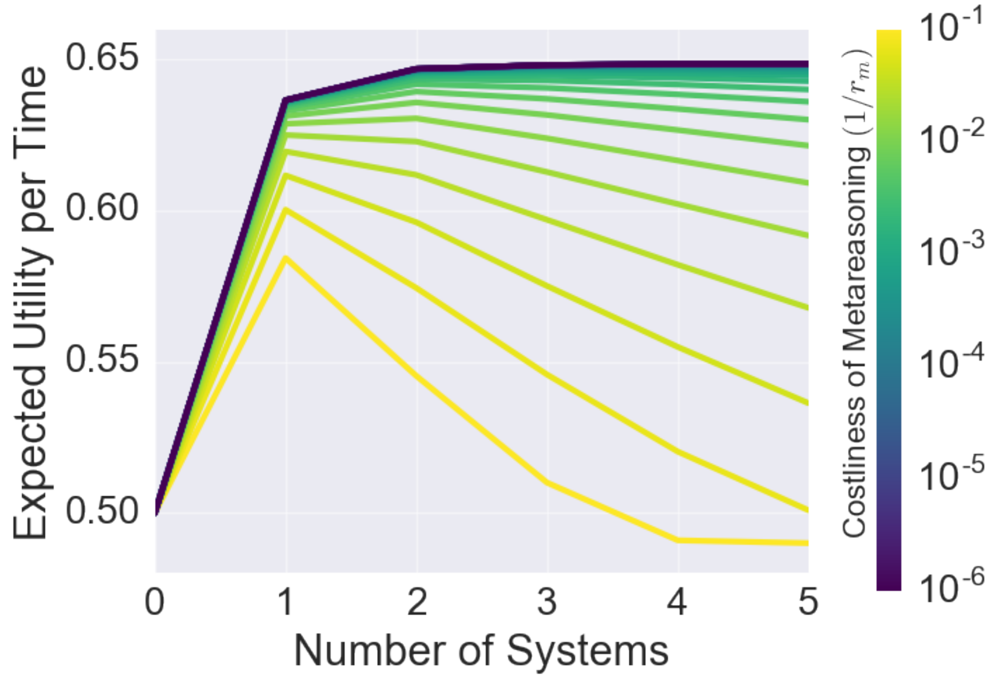

Figure 1. The reward rate in two-alternative forced choice (Simulation 1) usually peaks for a moderately small number of decision systems. The expected utility per time of the optimal choice of systems, $\mathcal{M}^\star$, as a function of the number of systems ($|\mathcal{M}|$). As the costliness of metareasoning, $\frac{1}{r_m}$ decreases, the optimal number of systems increases. In this example $\mathbb{E}[r_e] = 100$ and $\sigma(r_e) = 100$.

variable environment should comprise multiple cognitive systems with different characteristics. But at the same time, the number of systems should not be too high, or else the time spent on deciding which system to use, the metareasoning cost, will be too high. In other words, we hypothesize that the number of systems will depend on a tradeoff between the variability of the environment and the metareasoning cost. Our simulations show that this is indeed the case.

Simulation 1: Two-Alternative Forced Choice

Our first simulation focuses on the widely-used two-alternative forced choice (2AFC) paradigm, in which a participant is forced to select between two options. For example, categorization experiments often require their participants to decide whether the presented item belongs to the category or not, and psychophysics experiments often require participants to judge whether two stimuli are the same or different. Even in simple laboratory settings, judgments made within a 2AFC task seem to stem from systematically different modes of thinking. Therefore, 2AFC tasks are a prime setting to start in evaluating our theory for dual process systems. But before describing the details of our 2AFC simulation, we first review evidence of dual-process accounts of behavior in the 2AFC paradigm.

A very basic binary choice task presents an animal with a lever that it can either press to obtain food or decide not to press (Dickinson, 1985). It has been shown that early on in this task rodents' choices are governed by a flexible brain system that will stop pressing the lever when the they no longer want the food. By contrast, after extensive training their choices are controlled by a different, inflexible brain system that will continue to press the lever even when the reward is devaluated by poisoning the food. Interestingly, these two systems are preserved in the human brain and the same phenomenon has been demonstrated in humans (Balleine & O'Doherty, 2010).

Another example of two-alternative forced-choice is the probability learning task where participants repeatedly choose between two options, the first of which yields a reward with probability p_1 and the second of which yields a reward with probability $p_2 = 1 - p_1$. It has been found that depending on the incentives people tend to make these choices in two radically different ways (Shanks, Tunney, & McCarthy, 2002): When the incentives are low then people tend to use a strategy that chooses option one with a frequency close to p_1 and option two with a frequency close to p_2 – which can be achieved very efficiently (Vul, Goodman, Griffiths, & Tenenbaum, 2014). By contrast, when the incentives are high then people employ a choice strategy that maximizes their earnings by almost always choosing the option that is more likely to be rewarded – which requires more computation (Vul et al., 2014).

The dual systems perspective on 2AFC leaves open the normative question: what set of systems is optimal for the agent to be equipped with? To answer this question, we apply the methodology described in the previous section to the problem of bounded-optimal binary-choice.

Methods

As in the 2AFC probability learning task used by Shanks et al. (2002), the agent receives a reward of $+1$ for picking the correct action and 0 for picking the incorrect action. An unboundedly rational agent would always pick the action with a higher probability of being correct. Yet, although simple in set-up, computing the probability of an action being correct generally requires complex inferences over many interconnected variables. For example, if the choice is between turning left onto the highway or turning right to smaller backroads, estimating the probability of which action will lead to less traffic may require knowledge of when rush hour is, whether there is a football game happening, and whether there are accidents in either direction.

To approximate these often intractable inferences people appear to perform probabilistic simulations of the outcomes, and the variability and biases of their predictions (Griffiths & Tenenbaum, 2006; Lieder, Griffiths, Huys, & Goodman, 2018a) and choices (Vul et al., 2014; Lieder, Griffiths, & Hsu, 2018) match those of efficient sampling algorithms. Previous work has therefore modeled people as bounded-optimal sample-based agents, which draw a number of samples from the distribution over correct actions and then picks the action that was sampled most frequently. (Vul et al., 2014; Griffiths et al., 2015). In line with the prior work, we too model the agent as being a sample-based agent, described formally below.

Let a_0 and a_1 be the actions available to the agent where a_1 has a probability θ of being the correct action and a_0 has a probability $1 - \theta$ of being correct. The probability θ that a_1 is correct varies across different environments, reflecting the fact that in some settings it is easier to tell which action is correct than others. For example, it is obvious between the choice of a two-month old tomato and a fresh orange that the more nutritious choice is the latter. In this case, it is clear that the fresh orange is correct with probability near one. On the other hand, it may be quite difficult to decide between whether to attend graduate school at two universities with similar programs. In this case, the difference between the probabilities of each being correct may be quite marginal, and both might have close to a 0.5 chance of being correct. We model the variability in the difficulty of this choice by assuming that θ is equally likely to be any value in the range

$(0.5, 1)$, i.e $\theta \sim P_\theta = \text{Unif}(0.5, 1)$. We consider the range $(0.5, 1)$ instead of $(0, 1)$ without loss of generality because we can always rename the actions so that a_0 is more likely to be correct than a_1.

To make a decision the sample-based agent draws some number of samples k from the distribution over correct actions, $i \sim \text{Bern}(\theta)$, and picks the action a_i that it sampled more. [1] If the agent always draws k samples before acting, then its expected utility across all environments is

$$\mathbb{E}_\theta[U|k] = \int_\theta [P(a_1 \text{ is correct}) \cdot P(\text{Agent picks } a_1 \mid k) + P(a_0 \text{ is correct}) \cdot P(\text{Agent picks } a_0 \mid k)] P_\theta(d\theta) \,. \tag{1}$$

See Appendix A for a detailed derivation of how to calculate the quantity in Equation 1. If there were no cost for samples, then the agent could take an infinite number of samples to ensure choosing the correct action. But this is, of course, impractical in the real world because drawing a sample takes time and time is limited. Vul et al. (2014) show how the optimal number of samples changes based on the cost of sampling in various 2AFC problems. They parameterize the cost of sampling as the ratio, r_e, between the time for acting and the execution time of taking 1 sample. Suppose acting takes one unit of time, then the amount of time it takes to draw k samples is k/r_e. The total amount of time the agent takes is $1 + k/r_e$. Thus, the optimal number of samples the agent should draw to maximize its expected utility per unit time is

$$k^* = \arg\max_{k \in \mathbb{N}_0} \frac{\mathbb{E}_\theta[U|k]}{1 + \frac{k}{r_e}}. \tag{2}$$

When the time it takes to generate a sample is at least one tenth of the time it takes to execute the action ($r_e \leq 10$), then the optimal number of samples is either zero or one. In general, the first sample provides the largest gain in decision quality and the returns diminish with every subsequent sample. The point where the gain in decision quality falls below the cost of sampling

[1] If there is a tie, then the agent picks either a_0 or a_1 with equal probability. However, for odd k, the agent's expected utility after drawing k samples, $\mathbb{E}_\theta[U|k]$, is equal to its expected utility after drawing $k+1$ samples, $\mathbb{E}_\theta[U|k+1]$. Thus, we can restrict ourselves to odd k where no ties are possible.

Table 1

The optimal set of cognitive systems ($\mathcal{M}$) for the 2AFC task of Simulation 1 as a function of the number of systems ($|\mathcal{M}|$) and the variability of the environment (Var(r_e)) for $\mathbb{E}[r_e] = 100$ and $r_m = 1000$.

$\lvert\mathcal{M}\rvert$	Var(r_e)		
	10^3	10^4	10^5
1	3	3	1
2	3, 5	1, 5	1, 7
3	3, 5, 7	1, 3, 7	1, 3, 9
4	1, 3, 5, 7*	1, 3, 5, 7	1, 3, 7, 13

(*) Any set of four systems that included 3, 5, 7 was optimal.

depends on the value of r_e. Since this value can differ drastically across environments, achieving a near-optimal tradeoff in all environments requires adjusting the number of samples. Even a simple heuristic-based metareasoner that adapts the number of samples it takes based on a few thresholds on r_e does better than one which always draws the same number of samples (Icard, 2014).

Here, we study an agent that chooses how many samples to draw by metareasoning over a finite subset $\mathcal{M}$ of all possible numbers of samples. Furthermore, we assume that the time spent metareasoning increases linearly with the number of systems. By analogy to Vul et al. (2014), we formalize the metareasoning cost in terms of the ratio r_m of the time it takes to act over the time it takes to predict the performance of a single system.

We can again calculate the total amount of time the agent spends in the problem, while now taking into account the time spent on metareasoning. Just as before, the agent spends one unit of time executing its action, and k/r_e units of time to draw k samples. But now, we also account for the time it takes the agent to predict performance of a system: $1/r_m$. The total amount of time it takes the agent to metareason, i.e predict the performance of all systems, is $|\mathcal{M}|/r_m$. Therefore, the total amount of time is $1 + \frac{\pi_{\mathcal{M}}(r_e)}{r_e} + \frac{|\mathcal{M}|}{r_m}$. We assume the agent picks the optimal number of

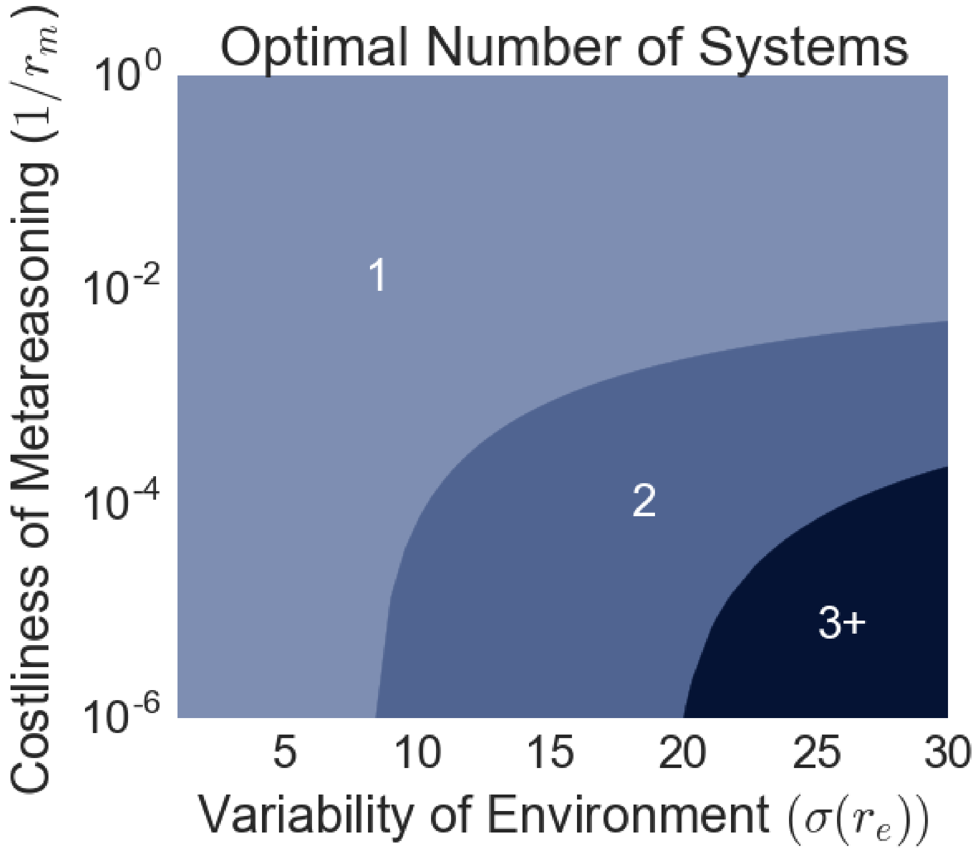

Figure 2. Performance of agents with different numbers of decision mechanisms in the 2AFC problem of Simulation 1. The plot shows the optimal number of decision systems as a function of the standard deviation of r_e and $1/r_m$. In this example $\mathbb{E}[r_e] = 10$.

samples out of the set of possible systems $\mathcal{M}$:

$$k^* = \arg\max_{k \in \mathcal{M} \cup \{0\}} \frac{\mathbb{E}_\theta[U|k]}{1 + \frac{k}{r_e} + \frac{|\mathcal{M}|}{r_m}}. \quad (3)$$

Given this formulation of the problem, we can now calculate the optimal set of systems for the agent. The set of cognitive systems that results in the optimal expected utility per time for the bounded sampling agent is

$$\mathcal{M}^* = \arg\max_{\mathcal{M} \subset \mathbb{N}} \mathbb{E}_{r_e}\left[\max_{k \in \mathcal{M} \cup \{0\}} \frac{\mathbb{E}_\theta[U|k]}{1 + \frac{k}{r_e} + \frac{|\mathcal{M}|}{r_m}}\right]. \quad (4)$$

Equation 4 resembles Equation 3 because both optimize the agent's expected utility per time. The difference is that Equation 3 calculates the optimal number of samples for a fixed cost of sampling, while Equation 4 calculates the optimal number of systems for a distribution of costs of sampling.

Note that the optimal set of systems depends on the distribution of the sampling cost r_e across different environments. Since sampling an action generally takes less time than executing the action, we assume that r_e is always greater than one. We can satisfy this constraint on r_e by modeling r_e as following a shifted Gamma distribution, i.e $r_e - 1 \sim \Gamma(\alpha, \beta)$.

Results

Figure 1 shows a representative example[2] of the expected utility per time as a function of the number of systems for different metareasoning costs. Under a large range of metareasoning costs the optimal number of systems is just one, but as the costliness of selecting a cognitive system decreases, the optimal number of systems increases. However even when the optimal number of systems is more than one, each additional system tends to only result in a marginal increase in utility, suggesting that one reason for few cognitive systems may be that the benefit of additional systems is very low.

Figure 2 shows that the optimal number of systems increases with the variance of r_e and decreases with the cost of selecting between cognitive systems (i.e., $\frac{1}{r_m}$). Interestingly, there is a large set of plausible combinations of variability and metareasoning cost for which the bounded-optimal agent has two cognitive systems. In addition, when the optimal number of systems is two, then the gap between the values of the two systems picked increases with the variance of r_e (see Table 1), resulting in one system that has high accuracy but high cost and another system that has low accuracy and low cost, which matches the characteristics of the systems posited by dual-process accounts. Thus, the conditions under which we would most expect to see two cognitive systems like the ones suggested by dual-process theories are when the environment is highly variable and arbitrating between cognitive systems is costly.

[2] For all experiments reported in this paper, we found that alternative values for $\mathbb{E}[r_e]$ or Var(r_e) did not change the qualitative conclusions, unless otherwise indicated.

Simulation 2: Sequential Decision-Making

Our first simulation modeled one-step decision problems in which the agent made a single choice between two options. In our second simulation, we turn to more complex, sequential decision problems, in which the agent needs to choose a sequence of actions over time in order to achieve its goal. In these problems, the best action to take at any given point depends on future outcomes and actions, thus leading to the need for *planning*. Furthermore, since actions only affect the environment probabilistically, it leads to the need for *planning under uncertainty*.

Although planning often allows us to make better decisions, planning places high demands on people's working memory and time (Kotovsky, Hayes, & Simon, 1985). This may be why research on problem solving has found that people use both planning and simple heuristics (Newell & Simon, 1972; Atwood & Polson, 1976; Kotovsky et al., 1985) and models of problem solving often assume that the mind is equipped with a planning strategy, such as means-ends analysis, and one or two simple heuristics such as hill-climbing (Newell & Simon, 1972; Gunzelmann & Anderson, 2003; Anderson, 1990).

Consistent with these findings, modern research on sequential decision-making points to the coexistence of two systems: a reflective, goal-directed system that uses a model of the environment to plan multiple steps into the future and a reflexive system that learns stimulus-response associations (Dolan & Dayan, 2013). Interestingly, people appear to select between these two systems in a manner consistent with rational metareasoning: When people are given a task where they can either plan two steps ahead to find the optimal path or perform almost equally well without planning, they often eschew planning (Daw et al., 2005; Kool, Cushman, & Gershman, 2016), but when the incentive structure is altered to make planning worthwhile then people predominantly rely on the planning system (Kool, Gershman, & Cushman, 2017). These findings are also consistent with Anderson's rational analysis of problem solving which assumed that people select between planning according to means-ends-analysis and a hill climbing heuristic according to a rational cost-benefit analysis (Anderson, 1990).

Working from the assumption that the mind is equipped with a planning-based system and a

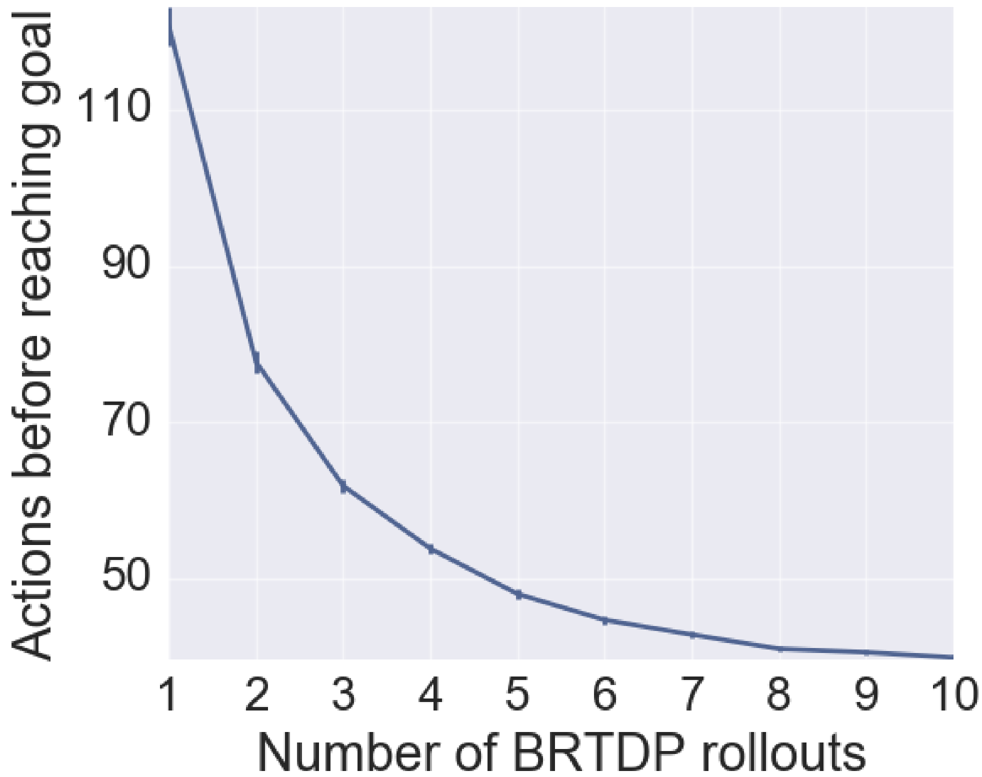

Figure 3. Performance of agents with different numbers of cognitive systems in planning under uncertainty (Simulation 2). The number of actions it takes an agent to reach a goal as a function of the number of simulated paths before each action. For 0 simulated paths the expected number of actions was 500 (the maximum allowed).

reflexive system, Daw et al. proposed a normative theory of how to choose which system to use (Daw et al., 2005). Here, we aim to derive a normative theory of what set of systems the mind should be equipped with in the first place.

Methods

Like Daw et al., we model the challenge of finding a sequence of actions that achieves the goal as a finite-horizon Markov decision problem (MDP; Sutton & Barto, 2018) with an absorbing goal-state. This type of MDP is formally defined by a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a cost function $c : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$ that measures how costly each action a is depending on the

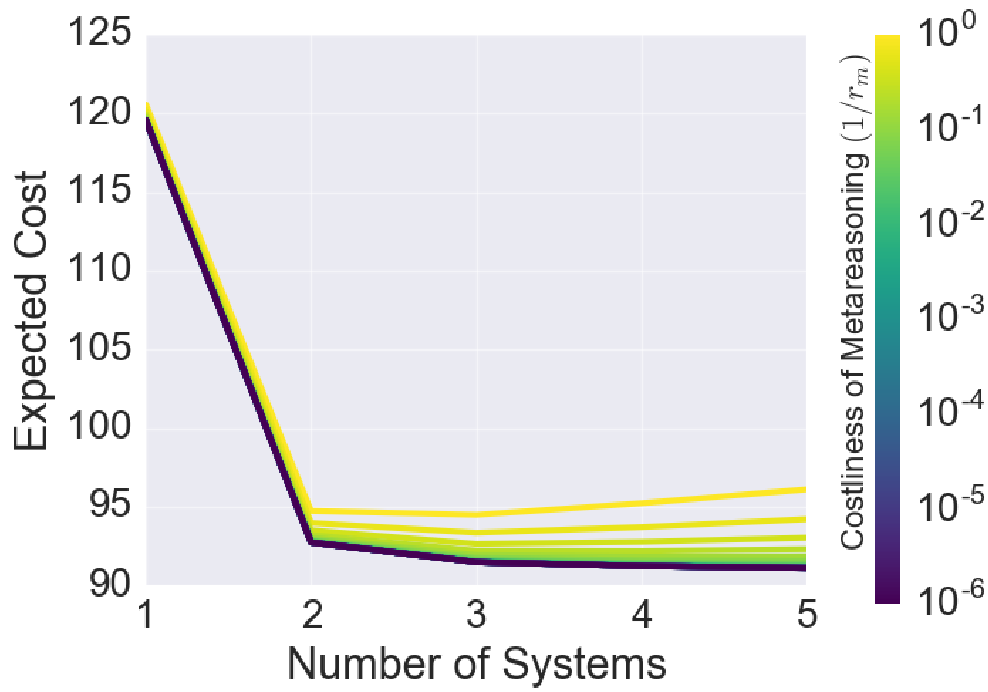

Figure 4. The expected cost incurred is a U-shaped function of the number of planning systems in Simulation 2. As the cost of selecting a planning system ($\frac{1}{r_m}$) decreases, the optimal number of systems increases. The expected cost of 0 systems was 500, thus 1 system provided the greatest reduction in cost. In this example $\mathbb{E}[r_e] = 100$, Var$(r_e) = 10^5$, and $c_a = 1$.

current state s, a transition probability model $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ that defines the probability of the next state given the current state and the action taken, an absorbing goal state g, and a time horizon h. Experience in these MDPs can be thought of as a set of trials or episodes. A trial ends once the agent reaches an absorbing goal-state g or it exceeds the maximal number of time steps allowed by the time horizon h.

In the standard formulation, at each time step, the agent takes an action, which depends upon its current state. The agent's action choices can be concisely represented by a policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$ that returns an action for each state. An optimal policy minimizes the expected sum of costs across the trial:

$$\pi^* = \arg\min_{\pi} \mathbb{E}\left[\sum_{i=0}^{N} c(s_i, \pi(s_i)) \,\middle|\, \pi\right], \tag{5}$$

where s_i is the state at time step i and N is the time step that the episode ends (either once the agent reaches the goal state g or the time horizon h is reached). The expectation is taken over the states at each time step, which are stochastic according to the transition model p.

However, this formulation of the problem ignores the fact that the agent needs to think to decide how to act, and that thinking also incurs cost. We extend the standard MDP formulation to account for the cost of thinking. At each time step, the agent has a thinking stage, followed by an acting stage. In the thinking stage, the agent executes a system t that (stochastically) decides on an action a. In the acting stage, the agent takes the action a. In addition to the cost $c(s, a)$ of acting, there is also a cost $f(t)$ that measures the cost of thinking with system t. Then, an optimal system minimizes the total expected cost of acting and thinking:

$$t^* = \arg\min_t \mathbb{E}\left[\sum_{i=0}^{N} c(s_i, a_i) + f(t) \,\middle|\, t\right], \tag{6}$$

where $a_0, \ldots a_N$ are the actions chosen by t at each time step and $s_0, \ldots s_N$ are the states at each time step. The expectation is taken over states and actions, which are stochastic because the transition model p and the system t are not necessarily deterministic.

The agent's thinking systems are based on *bounded real-time dynamic programming* (BRTDP; McMahan, Likhachev, and Gordan, 2005), a planning algorithm from the artificial intelligence literature. BRTDP simulates potential action sequences, and then uses these simulations to estimate an upper bound and lower bound on how good each action in each possible state. It starts with a heuristic bound, and then continuously improves the accuracy of its estimates. Depending on the number of simulations chosen, it can be executed for an arbitrarily short or long amount of time. Fewer simulations result in faster but less accurate solutions, while more simulations results in slower but more accurate solutions, making BRTDP particularly well-suited for studying metareasoning (Lin et al., 2015).

During the thinking stage, the agent chooses the number of action sequences to simulate (k), and then based on this simulations, uses BRTDP to update its estimate of how good each action is in each possible state. During the acting stage, the agent takes the action with the highest upper bound on its value. Thus the agent's policy is defined entirely by k, the number of action sequences it simulates. This type of policy corresponds to the Think*Act policy from Lin et al.

We consider environments in which there is a constant cost per action (c_a) from all non-goal states: $c(s, a) = c_a$. The cost of executing a system is linear in the number of simulated action sequences (k): $f(k) = c_e \cdot k$, where c_e is the cost of each mental simulation. We reparameterize the costs by the ratio of the cost of acting over the cost of thinking, $r_e = \frac{c_a}{c_e}$. Having defined the agent policy and the quotes, Equation 6 simplifies to

$$k^* = \arg\min_{k \in \mathbb{N}_0} \left(1 + \frac{k}{r_e}\right) \mathbb{E}\left[N|k\right], \tag{7}$$

where N is the number of time steps until the trial ends, either by reaching the goal state or the time horizon. See Appendix B for a derivation.

Equation 7 defines the optimal system for the agent to use for a particular decision problem, but we seek to investigate what set of systems is optimal for the agent to be equipped with for a range of decision problems. We assume that there is a distribution of MDPs the agent may encounter, and while r_e is constant within each problem, it varies across different problems. Therefore, optimally allocating finite computational resources requires metareasoning. We assume that metareasoning incurs a cost that is linear in the number of systems: $c_m \cdot |\mathcal{M}|$, where c_m is the cost required to predict the performance of a single system. Similarly we can reparametrize this cost using $r_m = c_a/c_m$, so that the cost of metareasoning becomes $|\mathcal{M}|/r_m$.

Assuming that the agent chooses optimally from its set of planning systems, the optimal set of systems that it should be equipped with is

$$\mathcal{M}^* = \arg\min_{\mathcal{M} \subset \mathbb{N}} \mathbb{E}_{r_e}\left[\min_{k \in \mathcal{M} \cup \{0\}} \left(1 + \frac{k}{r_e}\right) \mathbb{E}\left[N|k\right]\right] + \frac{|\mathcal{M}|}{r_m}. \tag{8}$$

We investigated the size and composition of the optimal set of planning systems for a simple 20×20 grid world where the agent's goal is to get from the lower left corner to the upper right corner with as little cost as possible. The horizon was set to 500, and the maximum number and length of simulated action sequences at any thinking stage were set to 10. BRTDP was initialized with a constant value function of 0 for the lower bound and a constant value function of 10^6 for the upper bound. This means that the agent's initial policy was to act randomly–which is highly suboptimal. For each environment, the ratio of the cost of action over the cost of planning

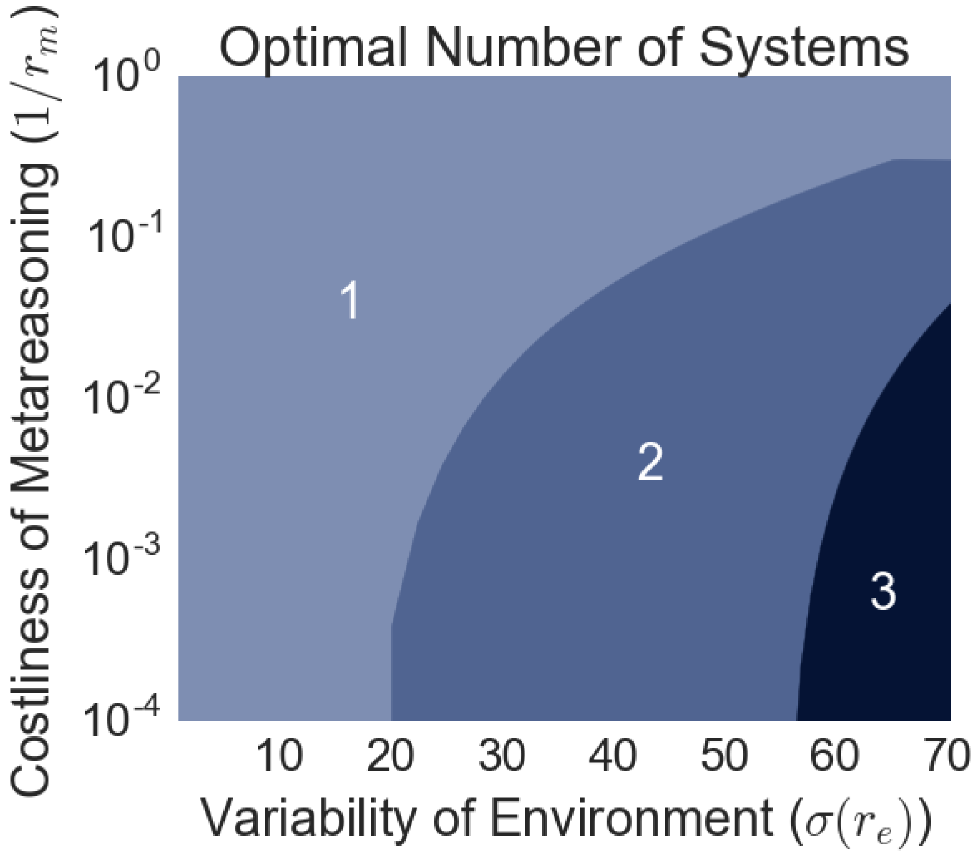

Figure 5. The optimal number of systems for planning under uncertainty (Simulation 2) as a function of the standard deviation of r_e and r_m for $\mathbb{E}[r_e] = 100$.

(r_e) was again drawn from a Gamma distribution and shifted by one, that is $r_e - 1 \sim \Gamma(\alpha, \beta)$. The expected number of steps required to achieve the goal $\mathbb{E}[N|k]$ was estimated via simulation (see Figure 3).

Results

We find that all our results match the 2-alternative forced choice setting extremely closely. Because the agent rarely reached the goal with zero planning ($\mathbb{E}[N|k = 0] = 500$) one system provided the largest reduction in expected cost with each additional system providing at most marginal reductions (Figure 4). The optimal number of systems increased with the variance of r_e and decreased with the metareasoning cost ($\frac{1}{r_m}$). This resulted in the optimal number of cognitive

Table 2

The optimal set of cognitive systems ($\mathcal{M}^\star$) for planning under uncertainty (Simulation 2) as a function of the number of systems ($|\mathcal{M}|$) and the variability of the environment (Var(r_e)) with $\mathbb{E}[r_e] = 100$.

$\|\mathcal{M}\|$	Var(r_e)		
	10^3	10^4	10^5
1	9	7	7
2	7, 9	4, 7	2, 7
3	1, 7, 9	4, 7, 9	1, 4, 9
4	1, 2, 7, 9	2, 4, 7, 9	1, 4, 7, 9

systems being two for a wide range of plausible combinations of variability and metareasoning cost (Figure 5). In addition, when the number of systems was two, the difference between the amount of planning performed by the two optimal systems increased with the variance of r_e.[3] This resulted in one system that does a high amount of planning but is costly and another system that plans very little but is computationally inexpensive, matching the characteristics of the two types of systems postulated by dual-process theories.

Simulation 3: Strategic interaction in a two-player game

Starting in the 1980s, researchers began applying dual-process theories to social cognition (Chaiken & Trope, 1999; Evans, 2008). One hypothesis for why the heuristic system exists is because exact logical or probabilistic reasoning is often computationally prohibitive. For instance, Herbert Simon famously argued that computational limitations place substantial constraints on

[3] This observation holds until the variance becomes extremely high ($\approx 10^7$ for Table 2), in which case both systems move towards lower values (Table 2). However, this is not a general problem but merely a quirk of the skewed distribution we used for r_e.

human reasoning (Simon, 1972, 1982). Such computational limitations become readily apparent in problems involving social cognition because the number of future possibilities explodes once the actions of others must be considered. For example, one of Simon's classic examples was chess, where reasoning out the best opening move is completely infeasible because it would require considering about 10^{120} possible continuations.

In this section, we show that our findings in decision-making and planning tasks about the optimal set of cognitive systems also applies to tasks that involve reasoning about decisions made by others. Specifically, we focus on strategic reasoning in Go, an ancient two-player game. Two-player games are the simplest and perhaps most widely used paradigm for studying strategic reasoning about other people's actions (Camerer, 2011). Although seemingly simple, it is typically impossible to exhaustively reason about all possibilities in a game, making heuristic reasoning necessary. This is especially true in Go, which has about 10^{360} continuations from the first move (compare this to chess which has "only" 10^{120} possible continuations).

Methods

We now describe the details of our simulation deriving bounded-optimal architectures for strategic reasoning in the game of Go.

The agent's thinking systems are based on a planning algorithm known as *Monte Carlo tree search* (MCTS) (Browne et al., 2012). Recently, AlphoGo, a computer system based on MCTS, became the first to defeat the Go world champion and achieve superhuman performance in the game of Go (Silver et al., 2016, 2017). Like other planning methods against adversarial opponents, MCTS works by constructing a game tree to plan future actions. Unlike other methods, MCTS selectively runs stochastic simulations (also known as *rollouts*) of different actions, rather than exhaustively searching through the entire game tree. In doing so, MCTS focuses on moves and positions whose values appear both promising and uncertain. In this regard, MCTS is similar to human reasoning (Newell & Simon, 1972).

Furthermore, the number of simulations used by MCTS affect how heuristic/accurate the

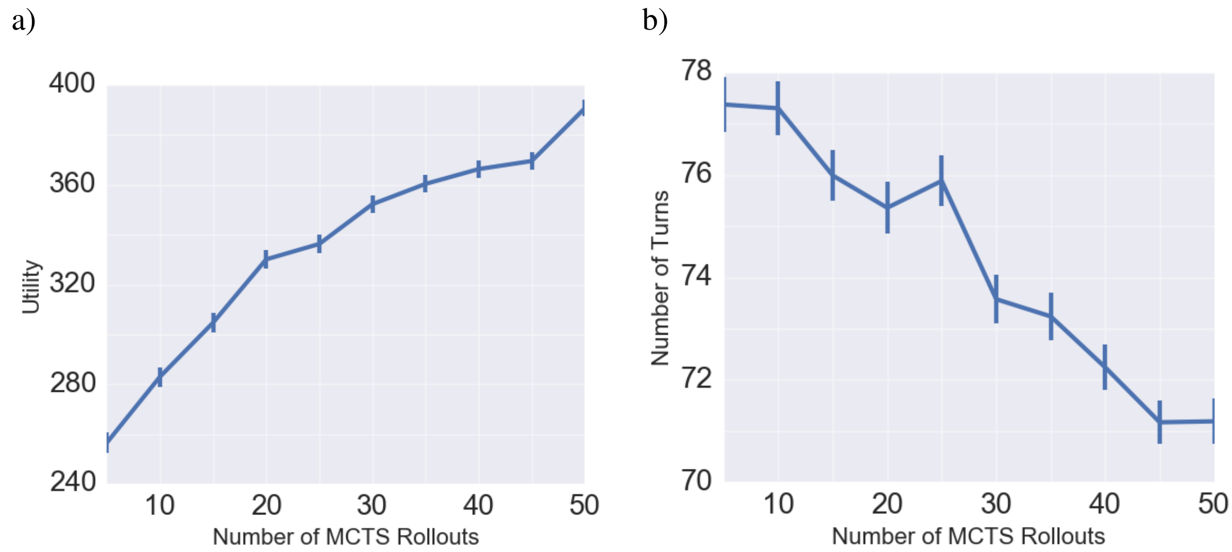

Figure 6. Performance as a function of the amount of reasoning in the game of Go (Simulation 3). As the amount of computation (number of simulations) increases, the likelihood of selecting a good action increases, thus resulting in larger utility (a) and the game tends to be won in increasingly fewer moves (b).

method is, making it well-suited for studying metareasoning. When the number of simulations is small, the algorithm is faster but less accurate. When the number of simulations is high, the algorithm is slower but more accurate. Thus, similar to the sequential decision making setting (Simulation 2), we assume that the agent metareasons over systems $\mathcal{M}$ that differ in how many simulations (k) to perform.

On each turn, there is a thinking stage and an acting stage. In the thinking stage, the agent executes a system that performs a number of stochastic simulations (k) of future moves and then updates its estimate of how good each action is, i.e. how likely it is to lead to a winning state. In the acting stage, the agent takes the action with the highest estimated value.

The agent attains a utility U based on whether it wins or loses the game. The unbounded agent would simply choose the number of simulations k that maximizes expected utility: $\mathbb{E}[U \mid k]$. However, the bounded agent incurs costs for acting and thinking. We assume that the cost for acting is constant: c_a. The cost for executing a system is linear in the number of simulations it performs: $k \cdot c_e$, where c_e is the cost of a single simulation. The bounded agent has to optimize a trade-off between its utility U and the costs of acting and thinking:

$$\mathbb{E}[U - (c_a + k \cdot c_e)N \mid k]\,, \tag{9}$$

where N is the number of turns until the game ends. For consistency, we can reparameterize this as $r_e = c_a/c_e$, the ratio between the cost of acting and the cost of thinking, and without loss of generality, we can let $c_a = 1$. Equation 9 then simplifies into

$$B(k, r_e) := \mathbb{E}\left[U - \left(1 + \frac{k}{r_e}\right) N \,\middle|\, k\right]\,. \tag{10}$$

The optimal system for the agent to choose given a fixed value of r_e is $k^*(r_e) = \arg\max_k B(k, r_e)$. The optimal set of cognitive systems $\mathcal{M}$ out of all possible systems $\mathcal{T}$ for strategic interaction is

$$\mathcal{M}^* = \underset{\mathcal{M} \subset \mathcal{T}}{\arg\max}\, \mathbb{E}\left[\max_k B(k, r_e)\right] - \frac{|\mathcal{M}|}{r_m}\,. \tag{11}$$

In this case, the expectation is taken over r_e, as the goal is to find the set of systems that is optimal *across* all problems in the environment.

In our simulations, the game is played on a 9×9 board. U is 500 if the agent wins, 250 if the game ends in a draw, and 0 if the agent loses. The opponent also runs MCTS with 5 simulations to decide its move. $\mathbb{E}[U \mid k]$ and $\mathbb{E}[N \mid k]$ are estimated using simulation (see Figure 6). For computational tractability, the possible number of simulations we consider are $\mathcal{T} = \{5, 10, \ldots, 50\}$.

Results

As in the previous tasks, the optimal number of systems depends on the variability of the environment and the difficulty of selecting between multiple systems (Figure 7). As the cost of metareasoning increases, the optimal number of systems decreases and the bounded-optimal agent comes to reason less and less. By contrast, the optimal number of systems increases with the variability of the environment. Furthermore, when the optimal number of systems is two, the

Table 3

The optimal set of cognitive systems ($\mathcal{M}^\star$) for strategic reasoning in the game of Go (Simulation 3) depending on the number of systems ($|\mathcal{M}|$) and the variability of the environment (Var(r_e)) for $\mathbb{E}[r_e] = 10$.

$\lvert\mathcal{M}\rvert$	Var(r_e)		
	10	10^2	10^3
1	10	10	10
2	10, 20	10, 20	10, 50
3	n/a*	10, 20, 50	10, 20, 50
4	n/a*	10, 20, 30, 50	10, 20, 30, 50

(*) This number of systems does not provide a noticeable increase in utility over fewer systems.

difference between the amount of reasoning performed by the two systems increases as the environment becomes more variable (Table 3). In conclusion, the findings presented in this section suggest that the kind of cognitive architecture that is bounded-optimal for simple decisions and planning (i.e., two systems with opposite speed-accuracy tradeoffs) is also optimal for reasoning about more complex problems, such as strategic interaction in games.

Simulation 4: Multi-alternative risky choice

Decision-making under risk is another domain in which dual-process theories abound (e.g., Steinberg, 2010; Mukherjee, 2010; Kahneman & Frederick, 2007; Figner, Mackinlay, Wilkening, & Weber, 2009), and the dual-process perspective was inspired in part by Kahneman and Tversky's ground-breaking research program on heuristics and biases (Kahneman, Slovic, & Tversky, 1982). Consistent with our resource-rational framework, previous research revealed that people make risky decisions by arbitrating between fast and slow decision strategies in an adaptive and flexible manner (Payne et al., 1993). When making decisions between the risky gambles shown in Figure 8 people adapt not only how much they think but also how they think

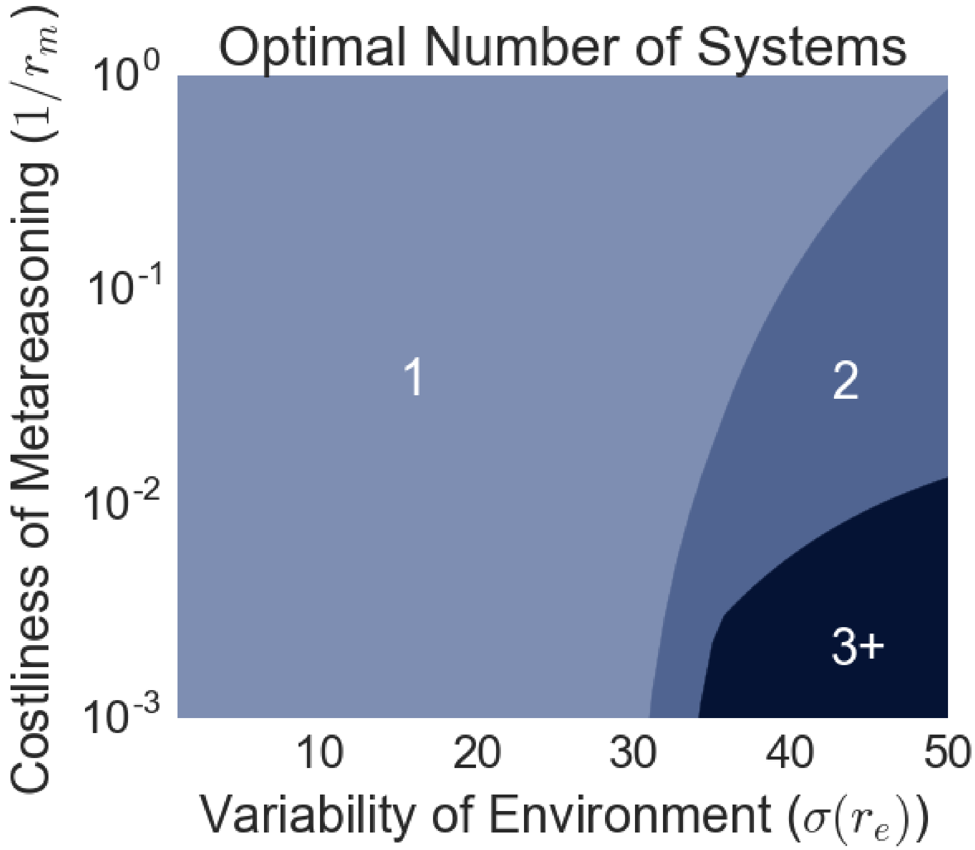

Figure 7. The optimal number of systems for strategic reasoning in the game of Go (Simulation 3) as a function of the standard deviation of r_e and $\frac{1}{r_m}$. $\mathbb{E}[r_e] = 100$ in this case.

about what to do. Concretely, people have been shown to use different strategies for different types of decision problems (Payne et al., 1988). For instance, when some outcomes are much more probably than others then people seem to rely on fast-and-frugal heuristics (Gigerenzer & Goldstein, 1996) like Take-The-Best which decides solely based on the most probably outcome that distinguishes between the alternatives and ignores all other possible outcomes. By contrast, when all outcomes are equally likely, people seem to integrate the payoffs for multiple outcomes into an estimate of the expected value of each gamble. Previous research has proposed at least ten different decision strategies that people might use when choosing between risky prospects (Payne et al., 1988; Thorngate, 1980; Gigerenzer & Selten, 2002). Yet, it has remained unclear how many decision strategies a single person would typically consider (Scheibehenne, Rieskamp, &

Wagenmakers, 2013). Here, we investigate how many decision strategies a boundedly optimal metareasoning agent should use in a multi-alternative risky-choice environment similar to the experiments by Payne et al.. Unlike in the previous simulations these strategies differ not only in how much computation they perform but also in which information they use and how they use it.

Figure 8. Illustration of the Mouselab paradigm used to study multi-alternative risky choice.

Methods

We investigated the size of the optimal subset of the ten decision strategies proposed by Payne et al. as a function of the metareasoning cost and the variability of the relative cost of reasoning. These strategies were the lexicographic heuristic (which corresponds to Take-The-Best), the semi-lexicographic heuristic, the weighted-additive strategy, choosing at random, the equal-weight heuristic, elimination by aspects, the maximum confirmatory dimensions heuristic, satisficing, and two combinations of elimination by aspects with the weighted additive strategy and the maximum confirmatory dimensions heuristic. Concretely, we determined the optimal number of decision strategies 5×30 environments that differed in the mean and the standard deviation of the distribution of r_e. The means were $10, 50, 100, 500$, and 1000, and the standard deviations were linearly spaced between 10^{-3} and 3 times the mean.

For each environment, four thousand decision problems were generated at random. Each problem presented the agent with the choice between five gambles with five possible outcomes. The payoffs for each outcome-gamble pair were drawn from a uniform distribution on the interval

$[0, 1000]$. The outcome probabilities differed randomly from problem to problem except that the second highest probability was always at most 25% of highest probability, the third highest probability was always at most 25% of the second-highest probability, and so on.

Based on previous work on how people select cognitive strategies (Lieder & Griffiths, 2017), our simulations assume that people generally select the decision-strategy that achieves the best possible speed-accuracy tradeoff. This strategy can be formally defined as the heuristic $s^\star$ with the highest value of computation (VOC; Lieder and Griffiths, 2017). Formally, for each decision problem d, an agent equipped with strategies $\mathcal{S}$ should choose the strategy

$$s^\star(d, \mathcal{S}, r_e) = \max_{s \in \mathcal{S}} \text{VOC}(s, d). \tag{12}$$

Following Lieder and Griffiths (2017) we define a strategy's VOC as decision quality minus decision cost. We measure the decision quality by the ratio of the expected utility of the chosen option over the expected utility of the best option, and we measure decision cost by the opportunity cost of the time required to execute the strategy. Formally, the VOC of making the decision d using the strategy s is

$$\text{VOC}(s, d) = \frac{\mathbb{E}\left[u(s(d))|d\right]}{\max_a \mathbb{E}\left[u(a)|d\right]} - \frac{1}{r_e} \cdot n_{\text{computations}}(s, d), \tag{13}$$

where $s(d)$ is the alternative that the strategy s chooses in the decision d, $\frac{1}{r_e}$ is the cost per decision operation, and $n_{\text{computations}}(s, d)$ is the number of cognitive operations it performs in this decision process. To determine the number of cognitive operations, we decomposed each strategy into a sequence of elementary information processing operations (Johnson & Payne, 1985) in the same way as Lieder and Griffiths (2017) did and counted how many of those operations each strategy performed on any given decision problem.

We estimated the optimal set of strategies,

$$\mathcal{S}^\star = \max_{\mathcal{S}} \mathbb{E}_{P(d)} \left[\text{VOC}(s^\star(d; \mathcal{S}, r_e), d) - \frac{1}{r_m} \cdot |\mathcal{S}| \right], \tag{14}$$

by approximating the expected value in Equation 14 by averaging the VOC over 4000 randomly generated decision problems. The resulting noisy estimates were smoothed with a Gaussian kernel with standard deviation 20. Then the optimal set of cognitive strategies was determined based on the smoothed VOC estimates for each combination of parameters. Finally, the number of strategies in the optimal sets was smoothed with a Gaussian kernel with standard deviation 10, and the smoothed values were rounded.

Results

As shown in Figure 9, we found that the optimal number of strategies increased with the variability of the environment and decreased with the metareasoning cost. Like in the previous simulations, the optimal number of decision systems increased from 1 for high metareasoning cost and low variability to 2 for moderate metareasoning cost and variability, and increased further with decreasing metareasoning cost and increasing variability. There was again a sizeable range of plausible values in which the optimal number of decision systems was 2. For extreme combinations of very low time cost and very high variability the optimal number of systems increased to up to 5. Although Figure 9 only shows the results for $\mathbb{E}[r_e] = 100$, the results for $\mathbb{E}[r_e] = 10$, 50, 500, and 1000 were qualitatively the same.

In this section, we applied our analysis to a more realistic setting than in the previous sections. It used psychologically plausible decision strategies that were proposed to explain human decision-making rather than algorithms. These strategies differed not only in how much reasoning they perform but also in how they reason about the problem. For this setting, where the environment comprised different kinds of problems favoring different strategies, one might expect that the optimal number of systems would be much larger than in the previous simulations. While we did find that having 3–5 systems became optimal for a larger range of metareasoning costs and variabilities, it is remarkable that having two systems was still bounded-optimal for a sizeable range of reasonable parameters. This finding suggests that our results might generalize to the much more complex problems people have to solve and people's much more sophisticated cognitive mechanisms.

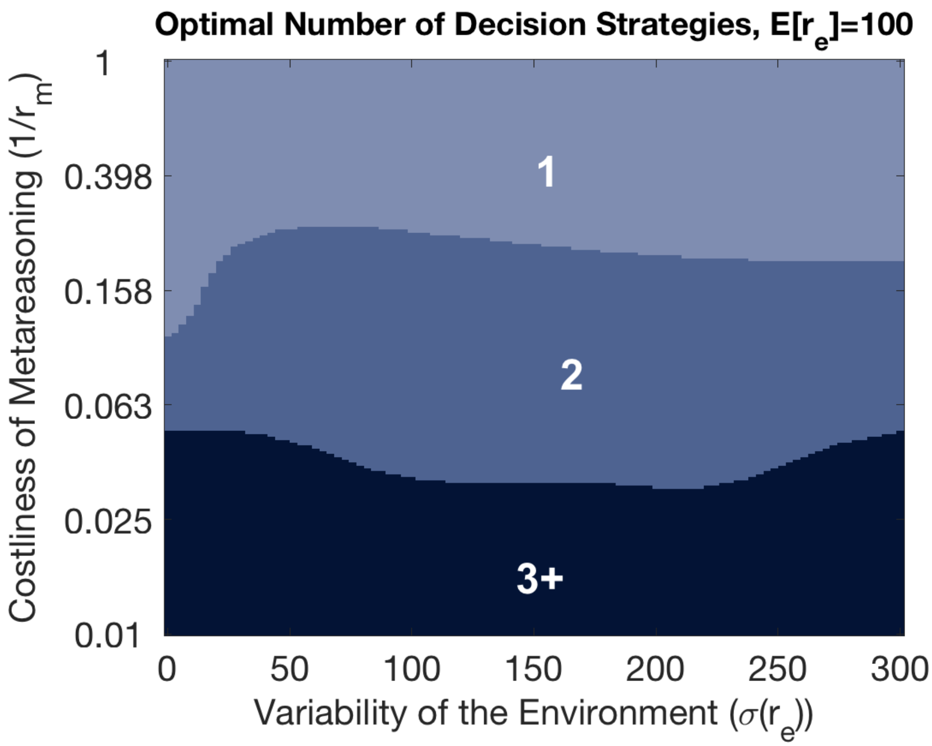

Figure 9. The optimal number of strategies for multi-alternative risky choice (Simulation 4) as a function of the standard deviation of r_e and r_m for $\mathbb{E}[r_e] = 100$.

General Discussion

We found that across four different tasks the optimal number and diversity of cognitive systems increases with the variability of the environment but decreases with the cost of predicting each system's performance. Each additional system tends to provide at most marginal improvements; so the optimal solutions tend to favor small numbers of cognitive systems, with two systems being optimal across a wide range of plausible values for metareasoning cost and variability. Furthermore, when the optimal number of cognitive systems was two, then these two systems tended to lie on two extremes in terms of time and accuracy. One of them was much faster but more error-prone whereas the second one was slower but more accurate. This might be why the human mind too appears to contain two opposite subsystems within itself – one that is fast but fallible and one that is slow but accurate. In other words, this mental architecture might have evolved to enable people to quickly adapt how they think and decide to the demands of different situations. Our analysis thereby provides a normative justification for dual-process theories.

The emerging connection between normative modeling and dual-process theories is remarkable because these approaches correspond to opposite poles in the debate about human rationality (Stanovich, 2011). In this debate, some researchers interpreted the existence of a fast, error-prone cognitive system whose heuristics violate the rules of logic, probability theory, and expected utility theory as a sign of human irrationality (Ariely, 2009; Marcus, 2009). By contrast, our analysis suggests that having a fast but fallible cognitive system in addition to a slow but accurate system may be the best possible solution. This implies that the variability, fallibility, and inconsistency of human judgment that result from people's switching between System 1 and System 2 should not be interpreted as evidence for human irrationality, because it might reflect the rational use of limited cognitive resources.

Limitations

One limitation of our analysis is that the cognitive systems we studied are simple algorithms that abstract away most of the complexity and sophistication of the human mind. A second limitation is that all of our tasks were drawn from the domains of decision-making and reasoning. However, our conclusion only depends on the plausible assumption that the cost of deciding which cognitive system to use increases with the number of systems. As long as this is the case, the optimal number of cognitive systems should still depend on the tradeoff between metareasoning cost and cognitive flexibility studied above, even though its exact value may be different. Thus, our key finding that the optimal number of systems increases with the variability of the environment and decreases with the metareasoning cost is likely to generalize to other tasks and the much more complex architecture of the human mind.

Third, our analysis assumed that the mind is divided into discrete cognitive systems to make the adaptive control over cognition tractable. While this makes selecting cognitive operations much more efficient, we cannot prove that it is bounded-optimal to approximate rational metareasoning in this way. Research in artificial intelligence suggests that there might be other ways to make metareasoning tractable. One alternative strategy is the meta-greedy approximation (Russell & Wefald, 1991a; Hay et al., 2012) which selects computations under the assumption that the agent will act immediately after executing the first computation. According to the directed cognition model (Gabaix & Laibson, 2005) this mechanism also governs the sequence of cognitive operations people employ to make economic decisions. This model predicts that people will always stop thinking when their decision cannot be improved by a single cognitive operation even when significant improvements could be achieved by a series of two or more cognitive operations. This makes us doubt that the meta-greedy heuristic would be sufficient to account for people's ability to efficiently solve complex problems, such as puzzles, where progress is often non-linear. This might be why when Gabaix, Laibson, Moloche, and Weinberg (2006) applied their model to multi-attribute decisions, they let it choose between macro-operators rather than individual computations. Interestingly, those macro-operators are similar to the cognitive systems studied here in that they perform different amounts of computation. Thus, the directed cognition model does not appear to eliminate the need for sub-systems but merely proposes a mechanism for how the mind might select and switch back-and-forth between them. Consistent with our analysis, the time and effort required by this mechanism increases linearly with the number of cognitive systems. While research in artificial intelligence as identified a few additional approximations to rational metareasoning, those are generally to specific computational processes and problems (Russell & Wefald, 1989; Lin et al., 2015; Vul et al., 2014) and would be applicable to only a small subset of people's cognitive abilities.

Relation to previous work

The work presented here continues the research programs of bounded rationality (Simon, 1956, 1982), rational analysis (Anderson, 1990) and resource-rational analysis (Griffiths et al., 2015) in seeking to understand how the mind is adapted to the structure of the environment and its limited computational resources. While previous work has applied the idea of bounded optimality to derive optimal cognitive strategies for an assumed cognitive architecture (Lewis et al., 2014;

Griffiths et al., 2015; Lieder, Griffiths, & Hsu, 2018; Lieder, Griffiths, Huys, & Goodman, 2018a) and the arbitration between assumed cognitive systems (Keramati et al., 2011), the work presented here derived the cognitive architecture itself. By suggesting that the human mind's cognitive architecture might be bounded-optimal our analysis complements and completes previous arguments suggesting that people make rational use of the cognitive architecture they are equipped with (Lewis et al., 2014; Griffiths et al., 2015; Lieder, Griffiths, & Hsu, 2018; Lieder, Griffiths, Huys, & Goodman, 2018a; Tsetsos et al., 2016; Howes et al., 2016). Taken together these arguments suggest that people might be resource-rational after all.

Conclusion and Future Directions

A conclusive answer to the question whether it is boundedly optimal for humans to have two types of cognitive systems will require more rigorous estimates of the variability of decision problems that people experience in their daily lives and precise measurements of how long it takes to predict the performance of a cognitive system. Regardless thereof, our analysis suggests that the incoherence in human reasoning and decision-making are qualitatively consistent with the rational use of a bounded-optimal set of cognitive systems rather than a sign of irrationality. Perhaps more importantly, the methodology we developed in this paper makes it possible to extend resource-rational analysis from cognitive strategies to cognitive architectures. This new line of research offers a way to elucidate how the architecture of the mind is shaped by the structure of the environment and the fundamental limits of the human brain.

Appendix A

2AFC

In this appendix, we derive the formula for the utility of making a decision based on k mental simulations used in our analysis of two alternative forced choice (i.e., Equation 1). Since there are two possible choices, there are two ways in which the agent can score a reward of 1, that is

$$\mathbb{E}_\theta[U|k] = \int_\theta [P(a_1 \text{ is correct}) \cdot P(\text{Agent picks } a_1 \mid k) \\ + P(a_0 \text{ is correct}) \cdot P(\text{Agent picks } a_0 \mid k)] P_\theta(d\theta) \,. \tag{1}$$

If a_i is the correct answer, then $i \sim \text{Bern}(\theta)$. The probability that the agent chooses a_i is equal to the probability that it sampled a_i more than $k/2$ times. The probability that the agent sampled a_0 more than $k/2$ times is $\Theta_{\text{CDF}}(k/2, \theta, k)$ where Θ_{CDF} is the binomial cumulative density function. Correspondingly, the probability that the agent sampled a_1 more than $k/2$ times is $1 - \Theta_{\text{CDF}}(k/2, \theta, k)$. Thus, we can write Equation 1 as

$$\mathbb{E}_\theta[U|k] = \int_\theta [\theta \left(1 - \Theta_{\text{CDF}}(k/2, \theta, k)\right) \\ + (1 - \theta) \left(\Theta_{\text{CDF}}\left(k/2, \theta, k\right)\right)] \, P_\theta(d\theta) \,.$$

Appendix B

Sequential Decision-Making

Here, we provide a derivation of how to simplify the expression for the optimal number of planning systems in Equation 6, that is

$$t^* = \arg\min_t \mathbb{E}\left[\sum_{i=0}^{N} c(s_i, a_i) + f(t) \,\middle|\, t\right], \tag{6}$$

to the expression in Equation 7, that is

$$k^* = \arg\min_{k \in \mathbb{N}_0} \left(1 + \frac{k}{r_e}\right) \mathbb{E}\left[N|k\right]. \tag{7}$$

Our reasoning behind this derivation is as follows: Since the cost of each thinking system is linear in the number of simulations, i.e. $c_e \cdot k$, we can replace $f(t)$ with $c_e \cdot k$ in the expectation in Equation 6. Since the cognitive systems are distinguished by the number of simulations they do, we can condition on the number of simulations k instead. Therefore, the expectation in Equation 6 becomes

$$\mathbb{E}\left[\sum_{i=0}^{N} c(s_i, a_i) + c_e \cdot k \,\middle|\, k\right].$$

The cost of acting from non-goal states is constant, i.e. $c(s_i, a_i) = c_a$. Therefore, the expectation simplifies to 6 becomes

$$\mathbb{E}\left[\sum_{i=0}^{N} c_a + c_e \cdot k \,\middle|\, k\right] = \mathbb{E}[N(c_a + c_e \cdot k) \mid k].$$

We can reparameterize using $r_e = c_a / c_e$ by substituting c_e with c_a / r_e:

$$\mathbb{E}\left[N\left(c_a + \frac{c_a}{r_e} \cdot k\right) \,\middle|\, k\right] = c_a \mathbb{E}\left[\left(1 + \frac{k}{r_e}\right) N \,\middle|\, k\right].$$

We now arrive at Equation 6 by picking the cognitive system (number of simulations) that minimizes the above quantity.

$$k^* = \arg\min_k c_a \mathbb{E}\left[\left(1 + \frac{k}{r_e}\right) N \,\middle|\, k\right] = \arg\min_k \mathbb{E}\left[\left(1 + \frac{k}{r_e}\right) N \,\middle|\, k\right].$$

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Psychology Press.

Ariely, D. (2009). *Predictably irrational*. New York: Harper Collins.

Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology*, *8*(2), 191–216.

Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science*, *35*(3), 499–526.

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69.

Bhui, R., & Gershman, S. J. (2017). Decision by sampling implements efficient coding of psychoeconomic functions. *bioRxiv*, 220277.

Boureau, Y.-L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in cognitive sciences*, *19*(11), 700–710.

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., … Colton, S. (2012). A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, *4*(1), 1–43.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in cognitive sciences*, *3*(2), 57–65.

Crockett, M. J. (2013). Models of morality. *Trends in cognitive sciences*, *17*(8), 363–366.

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and social psychology review*, *17*(3), 273–292.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, *8*(12), 1704–1711.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135–168.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society B*, *308*(1135), 67–78.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312 - 325.

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in cognitive sciences*, *7*(10), 454–459.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition. *Perspectives on Psychological Science*, *8*(3), 223-241.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia card task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 709.

Gabaix, X., & Laibson, D. (2005). *Bounded rationality and directed cognition* (Tech. Rep.). Cambridge, MA: Harvard University.

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, *96*(4), 1043–1068.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European review of social psychology*, *2*(1), 83–115.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, *103*(4), 650.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Greene, J. D. (2015). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *The Law & Ethics of Human Rights*, *9*(2), 141–172.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, *7*(2), 217–229.

Griffiths, T. L., & Tenenbaum, J. B. (2001). Randomness and coincidences: Reconciling intuition and probability theory. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 370–375).

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, *17*(9), 767–773.

Gunzelmann, G., & Anderson, J. R. (2003). Problem solving: Increased planning with practice. *Cognitive systems research*, *4*(1), 57–76.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychological review*, *114*(3), 704.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: why three heads are better than four. *Psychological review*, *116*(2), 454.

Hay, N., Russell, S. J., Tolpin, D., & Shimony, S. (2012). Selecting Computations: Theory and Applications. In N. de Freitas & K. Murphy (Eds.), *Proceedings of the 28th conference on uncertainty in artificial intelligence.* Corvallis: AUAI Press.

Horvitz, E. J., Cooper, G. F., & Heckerman, D. E. (1989). Reflection and action under scarce resources: Theoretical principles and empirical study. In *Proceedings of the eleventh international joint conference on artificial intelligence* (pp. 1121–1127). San Mateo, CA: Morgan Kaufmann.

Horvitz, E. J., & Rutledge, G. (1991). Time-dependent utility and action under uncertainty. In

Proceedings of the seventh conference on uncertainty in artificial intelligence (pp. 151–158).

Howes, A., Warren, P. A., Farmer, G., El-Deredy, W., & Lewis, R. L. (2016). Why contextual preference reversals maximize expected value. *Psychological review*, *123*(4), 368.

Icard, T. (2014). Toward boundedly rational analysis. In *Proceedings of the 36th annual conference of the cognitive science society* (pp. 637–642).

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management science*, *31*(4), 395–414.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Strauss and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment.* Cambridge, UK: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Frederick, S. (2007). Frames and brains: Elicitation and control of response tendencies. *Trends in cognitive sciences*, *11*(2), 45–46.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582 -59.

Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology*, *7*(5), e1002055.

Khaw, M. W., Li, Z., & Woodford, M. (2017). *Risk aversion as a perceptual bias* (Tech. Rep.).

Cambridge, MA: National Bureau of Economic Research.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLoS computational biology*, *12*(8), e1005090.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological science*, *28*(9), 1321–1333.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? evidence from tower of hanoi. *Cognitive psychology*, *17*(2), 248–294.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in cognitive science*, *6*(2), 279–311.

Lieder, F., & Griffiths, T. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*, 762 -794.

Lieder, F., & Griffiths, T. (in revision). *Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources.*

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, *125*(1), 1.

Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (2018a). The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review*, *25*(1), 322–349.

Lieder, F., Griffiths, T. L., Huys, Q. J. M., & Goodman, N. D. (2018b). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review*, *25*(2), 775–784.

Lieder, F., Krueger, P. M., & Griffiths, T. L. (2017). An automatic method for discovering rational heuristics for risky choice. In *Proceedings of the 39th annual meeting of the cognitive science society. austin: Cognitive science soc.*

Lieder, F., Plunkett, D., Hamrick, J. B., Russell, S. J., Hay, N. J., & Griffiths, T. L. (2014). Algorithm selection by rational metareasoning as a model of human strategy selection. In *Advances in neural information processing systems* (Vol. 27).

Lin, C. H., Kolobov, A., Kamar, E., & Horvitz, E. J. (2015). Metareasoning for planning under uncertainty. In *Proceedings of the 24th international conference on artificial intelligence* (pp. 1601–1609). AAAI Press.

Marcus, G. (2009). *Kluge: The haphazard evolution of the human mind*. Boston: Houghton Mifflin Harcourt.

McMahan, H. B., Likhachev, M., & Gordon, G. J. (2005). Bounded real-time dynamic programming: RTDP with monotone upper bounds and performance guarantees. In *Proceedings of the 22nd international conference on machine learning* (pp. 569–576).

Milli, S., Lieder, F., & Griffiths, T. L. (2017). When does bounded-optimal metareasoning favor few cognitive systems? In *Proceedings of the thirty-first AAAI conference on artificial intelligence* (pp. 4422–4428).

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological review*, *117*(1), 243.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104) (No. 9). Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum Press.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Parpart, P., Jones, M., & Love, B. (2017). Heuristics as Bayesian inference under extreme priors. *Cognitive Psychology*.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.

Russell, S. J., & Subramanian, D. (1995). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, *2*, 575–609.

Russell, S. J., & Wefald, E. (1989). On optimal game-tree search using rational meta-reasoning. In *Proceedings of the 11th international joint conference on artificial intelligence-volume 1* (pp. 334–340).

Russell, S. J., & Wefald, E. (1991a). *Do the right thing: studies in limited rationality*. Cambridge, MA: MIT press.

Russell, S. J., & Wefald, E. (1991b). Principles of metareasoning. *Artificial intelligence*, *49*(1-3), 361–395.

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*, *120*(1), 39.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*(40), 99-124.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*, 484-489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129-138.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, *1*(1), 161–176.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason*. Cambridge, MA: MIT press.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665–690.

Stanovich, K. E. (2009). *Decision making and rationality in the modern world*. Oxford, UK: Oxford University Press.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford, UK: Oxford University Press.

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental psychobiology*, *52*(3), 216–224.

Sutherland, S. (2013). *Irrationality: The enemy within*. London, UK: Pinter & Martin Ltd.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). The rational basis of representativeness. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 1036–41).

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219–225.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, *113*(11), 3102–3107.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

van der Meer, M., Kurth-Nelson, Z., & Redish, A. D. (2012). Information processing in decision-making systems. *The Neuroscientist*, *18*(4), 342–359.

Von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior*. Princeton, NJ: Princeton university press.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive science*, *38*(4), 599–637.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly journal of experimental psychology*, *20*(3), 273–281.