When Does Bounded-Optimal Metareasoning Favor Few Cognitive Systems?

Smitha Milli
EECS Department
University of California
Berkeley, CA 94720
smilli@berkeley.edu

Falk Lieder
Helen Wills Neuroscience Institute
University of California
Berkeley, CA 94720
falk.lieder@berkeley.edu

Thomas L. Griffiths
Department of Psychology
University of California
Berkeley, CA 94720
tom_griffiths@berkeley.edu

Abstract

While optimal metareasoning is notoriously intractable, humans are nonetheless able to adaptively allocate their computational resources. A possible approximation that humans may use to do this is to only metareason over a finite set of cognitive systems that perform variable amounts of computation. The highly influential "dual-process" accounts of human cognition, which postulate the coexistence of a slow accurate system with a fast error-prone system, can be seen as a special case of this approximation. This raises two questions: how many cognitive systems should a bounded optimal agent be equipped with and what characteristics should those systems have? We investigate these questions in two settings: a one-shot decision between two alternatives, and planning under uncertainty in a Markov decision process. We find that the optimal number of systems depends on the variability of the environment and the costliness of metareasoning. Consistent with dual-process theories, we also find that when having two systems is optimal, then the first system is fast but error-prone and the second system is slow but accurate.

Introduction

An agent should not think long about what to do when it is about to get hit by a car, but should definitely do so before declaring war. The optimal amount to think before making a decision is dependent on the setting, involving a trade-off between the cost of thinking and the cost of making an error. Thus, in order for an agent to be optimal across different settings it must *metareason* to adaptively choose the correct amount to think (Russell and Wefald 1991a).

Optimal metareasoning is generally computationally intractable (Russell and Wefald 1991a; Lin et al. 2015). Despite this, humans are still able to adaptively allocate their computational resources in inference and decision-making (Vul et al. 2014; Griffiths, Lieder, and Goodman 2015; Gershman, Horvitz, and Tenenbaum 2015). One method by which humans may perform approximate metareasoning is to choose between a discrete set of cognitive systems that perform variable amounts of computation (Lieder et al. 2014; Lieder and Griffiths 2015), rather than directly determining the exact amount of computation that should be done. This kind of approximation is a generalization of dual-process accounts of human cognition (Evans 2008; Kahneman 2011), which posit the existence of two types of cognitive systems, one that is slow and accurate and another that is fast but error-prone.

The postulated existence of distinct cognitive systems raises two interesting questions. First, how many systems should an agent have? Second, what characteristics should those systems have? These questions are important for understanding what the existence of multiple cognitive systems entails for human rationality (Evans 2008; Stanovich 2011) as well as for designing bounded-optimal AI systems (Russell and Subramanian 1995) that can metareason efficiently.

We investigate these questions in two settings. The first setting is a simple two-alternative forced choice task where the agent samples from a posterior distribution to infer which of two options is better. In this case, metareasoning selects how many samples to draw before taking an action (Vul et al. 2014). The second setting involves planning under uncertainty in a Markov decision process (MDP) (Lin et al. 2015). In this case, metareasoning determines the number of rollouts to perform to approximate the solution to the MDP with bounded real-time dynamic programming (BRTDP; McMahan, Likhachev, and Gordan, 2005).

Our experiments show that the optimal number of systems increases with the variability of the environment but decreases with the costliness of metareasoning. 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) suggesting that bounded optimality might provide a rational reinterpretation of those theories.

Background

We draw on the idea that the mind is composed of multiple cognitive systems, as instantiated in dual-process theories of human cognition, and the ideas of bounded optimality and rational metareasoning developed in artificial intelligence research. We review these ideas in turn.

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. This idea has subsequently 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 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 and Dayan 2013). The mechanisms employed by these systems have been mapped onto model-based versus model-free reinforcement learning algorithms that are used to solve Markov decision processes. A model-free and model-based distinction has also been suggested to account for the nature of the two systems posited to underlie moral reasoning (Cushman 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, and Tenenbaum 2015; Griffiths, Lieder, and Goodman 2015), we ask whether there are any principled reasons why bounded agents should be equipped with two types of reasoning systems rather than just one system, or three or more systems.

Metareasoning

Intelligent agents with a large repertoire of possible computations can solve problems in many different ways. On the one hand, this flexibility enables highly efficient solutions that exploit the structure of each problem. On the other hand, this flexibility poses the challenging problem of deciding when to perform which computation. This problem is known as rational metareasoning (Russell and Wefald 1991a). Inspired by psychology, many AI architectures solve this problem through metacognition (Anderson and Oates 2007; Cox and Raja 2011) which includes the monitoring and control of computation as well as metalearning (Smith-Miles 2009; Schaul and Schmidhuber 2010; Thornton et al. 2013).

The adaptive control of computation is critical for intelligent systems to be able to solve complex and potentially time-critical problems on performance-limited hardware (Horvitz, Cooper, and Heckerman 1989; Russell and Wefald 1991a). For instance, it is necessary for a patient-monitoring system used in emergency medicine to metar-eason in order to decide when to terminate diagnostic reasoning and recommend treatment (Horvitz and Rutledge 1991). Selecting computations can be formalized as a meta-level Markov decision process (Hay et al. 2012). This formulation highlights that selecting computations optimally is a computation-intensive problem 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 computation (Lin et al. 2015; Hay et al. 2012; Russell and Wefald 1991b). Yet, people appear capable of allocating their finite computational resources near-optimally with very little effort (Gershman, Horvitz, and Tenenbaum 2015; Keramati, Dezfouli, and Piray 2011).

To make optimal use of their finite computational resources *bounded-optimal* agents (Russell and Subramanian 1995) must optimally distribute their resources between metareasoning and reasoning about the world. This generally requires an efficient approximation to optimal metareasoning. But exactly how rational metareasoning should be approximated to maximize the agent's overall computational efficiency is an open problem. One common approximation to optimal metareasoning is the meta-myopic approximation (Russell and Wefald 1991b; Hay et al. 2012) which assumes that the agent will act immediately after executing the first computation. This approximation fails when no single computation can improve the agent's decision but a longer sequence of computations would improve it significantly.

Another strategy to approximate rational metareasoning is to exploit computational properties of the specific computational process being controlled (Russell and Wefald 1989; Lin et al. 2015; Vul et al. 2014). For instance, Vul et al. exploited the statistical properties of sampling from a Bernoulli distribution to solve the problem of when to terminate deliberation using the sequential probability ratio test. Lin et al. were also able to make metareasoning about planning tractable by making a meta-myopic assumption and then exploiting specific convergence properties of BRTDP.

A third approximation strategy is to metareason only over a limited number of computational mechanisms that generate decisions (Lieder et al. 2014). This approximation can drastically reduce the computational complexity of metareasoning while achieving human-level performance (Lieder et al. 2014; Lieder and Griffiths 2015). However, 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 induces a tradeoff between the efficiency of metareasoning and the performance of the best available computational mechanism. Although expanding the computational repertoire increases the expected performance of the selected mechanism, it also makes metareasoning more costly. 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.

Optimal Choice of Cognitive Systems

At an abstract level, each cognitive system can be characterized by how it trades off speed versus accuracy. In the remainder of this paper we characterize the optimal set of cognitive systems an agent should be equipped with in terms of the systems' speed-accuracy tradeoffs. To address this question as straight-forwardly as possible, we study sets of simple decision-making algorithms that abstract away the complexity of real cognitive systems but capture the diversity of their speed-accuracy tradeoffs. Specifically, we investigate what determines the optimal set of cognitive systems $\mathcal{M}^\star$ for an agent acting in an environment characterized by a distribution P over a set of decision problems D. An unboundedly rational agent would always pick the utility-maximizing action, $a^* = \arg\max_{a \in \mathcal{A}} U_d(a)$, where $U_d(a)$ is the utility of taking action a in decision problem d and $\mathcal{A}$ is the set of possible actions. However, any physical agent is necessarily bounded. This means when the agent executes an algorithm $t \in \mathcal{T}$ to decide which action to take it incurs some execution cost $c_e(t)$ corresponding to the required time and computational resources. Hence, bounded agents have to balance the cost of thinking with the expected value of the subsequent action A. In this case, the optimal thinking algorithm, t^*, for the agent to use on decision problem d is:

$$t^* = \arg\max_{t \in \mathcal{T}} \mathbb{E}\left[U_d(A)|t\right] - c_e(t). \tag{1}$$

To be optimal across all decision problems in its environment, the agent must metareason to adapt its algorithm t^* to each individual problem $d \in D$. However, $\mathcal{T}$ is a potentially infinite set. Hence, metareasoning over all algorithms in $\mathcal{T}$ can be intractable or prohibitively expensive. For this reason, instead of forcing an agent to metareason over the entirety of $\mathcal{T}$, the designer of the agent may opt to equip the agent only with a small set of cognitive systems $\mathcal{M} \subset \mathcal{T}$ to make metareasoning tractable.

However as the number of systems increases, so does the difficulty of choosing between them. To capture this, we introduce the metareasoning cost $c_m : \mathcal{P}(\mathcal{T}) \to \mathbb{R}$. The optimal $\mathcal{M}^*$ for the agent designer to choose is

$$\mathcal{M}^* = \arg\max_{\mathcal{M} \subset \mathcal{T}} \mathbb{E}\left[U_D(A) - c_e(\pi_{\mathcal{M}}(D))\right] - c_m(\mathcal{M}), \tag{2}$$

where $\pi_{\mathcal{M}}(D) : \mathcal{D} \to \mathcal{M}$ is the agent's metareasoning policy, the way the agent chooses which system to pick from the set $\mathcal{M}$, and A is the action selected after using that system. Next, we evaluate $\mathcal{M}^*$ through computational experiments in two settings: a simple two-alternative forced choice task and planning under uncertainty.

Two-Alternative Forced Choice

Our first experiment focuses on a two-alternative forced choice (2AFC) task, a one-shot decision problem where the agent must choose between two actions. Here, the agent gets 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
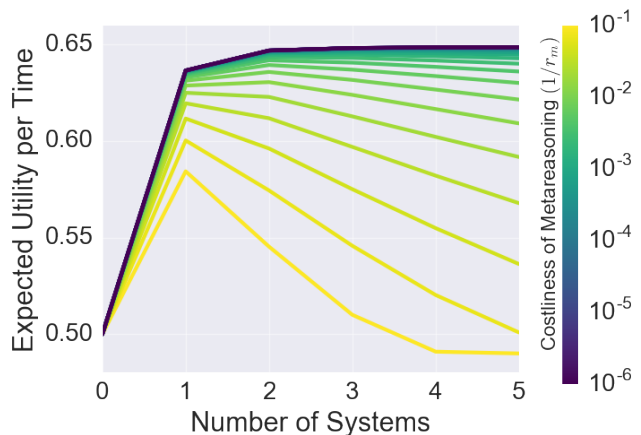

Figure 1: The expected utility per time of the optimal choice of systems, $\mathcal{M}$, as $|\mathcal{M}|$ increases. As the costliness of metareasoning, $\frac{1}{r_m}$ decreases, the optimal number of systems increases. In this example $\mathbb{E}[r_e] = 100$ and $\sigma(r_e) = 100$.

$|\mathcal{M}|$	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.

Table 1: The optimal set of algorithms, $\mathcal{M}$, for different values of $|\mathcal{M}|$ as Var(r_e) increases. This example is the 2AFC setting with $\mathbb{E}[r_e] = 100$ and $r_m = 1000$.

complex inferences over many interconnected variables. To approximate these often intractable inferences people appear to perform probabilistic simulations of the outcomes, and the variability and biases of their predictions (Griffiths and Tenenbaum 2006; Lieder, Griffiths, and Goodman 2012) and choices (Vul et al. 2014; Lieder, Hsu, and Griffiths 2014) match those of efficient sampling algorithms. Previous work has therefore modeled people as bounded-optimal sample-based agents (Vul et al. 2014; Griffiths, Lieder, and Goodman 2015). A sample-based agent draws a number of samples from the distribution over correct actions and then picks the action that was sampled most frequently.

For the 2AFC setting, let a_0 and a_1 be the actions available to the agent. We assume that the correct action a_i given all of the information available to the agent was sampled from a Bernoulli distribution, that is $i \sim \text{Bern}(\theta)$. The probability θ that a_1 is correct varies across different problems. We model this by sampling θ from a uniform distribution. Without loss of generality, we assume that $\theta \sim P_\theta = \text{Unif}(0.5, 1)$ because we can always rename the actions so that a_1 is more likely to be correct than a_0.

Under these assumptions, the expected value of the utility

U of a decision made based on k samples is given by

$$\mathbb{E}_\theta[U|k] = \int_\theta (\theta\,(1 - \Theta_{\text{CDF}}(k/2, \theta, k)))$$
$$+ (1 - \theta)\,(\Theta_{\text{CDF}}(k/2, \theta, k)))\,P_\theta(d\theta), \quad (3)$$

where Θ_{CDF} is the binomial cumulative density function.

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. In this setting, the optimal number of samples an agent should draw to maximize its expected utility per unit time is

$$k^* = \arg\max_{k \in \mathbb{N}_0} \frac{\mathbb{E}_\theta[U|k]}{1 + \frac{k}{r_e}}. \quad (4)$$

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 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. This allows us to formalize the metareasoning problem as computing a metacognitive policy $\pi_\mathcal{M} : \mathbb{R} \to \mathcal{M} \cup \{0\}$ that decides which of the agent's sampling systems ($\mathcal{M}$) to use given the relative cost of sampling (r_e). Given this formulation of the problem, the optimal set of systems that a bounded sampling agent should be equipped with is

$$\mathcal{M}^* = \arg\max_{\mathcal{M} \subset \mathbb{N}} \mathbb{E}_{r_e}\left[\frac{\mathbb{E}_\theta[U|\pi_\mathcal{M}(r_e)]}{1 + \frac{\pi_\mathcal{M}(r_e)}{r_e} + \frac{|\mathcal{M}|}{r_m}}\right], \quad (5)$$

where $\pi_\mathcal{M}(r_e)$ can be an arbitrary metareasoning policy. In the remainder of this paper, we will focus on bounded agents that metareason optimally according Equation (4). For these agents the optimal set of systems defined in Equation 5 becomes

$$\mathcal{M}^* = \arg\max_{\mathcal{M} \subset \mathbb{N}} \mathbb{E}_{r_e}\left[\max_{k \in \mathcal{M} \cup \{0\}} \frac{\mathbb{E}_\theta[U|k]}{1 + \frac{k}{r_e} + \frac{|\mathcal{M}|}{r_m}}\right]. \quad (6)$$
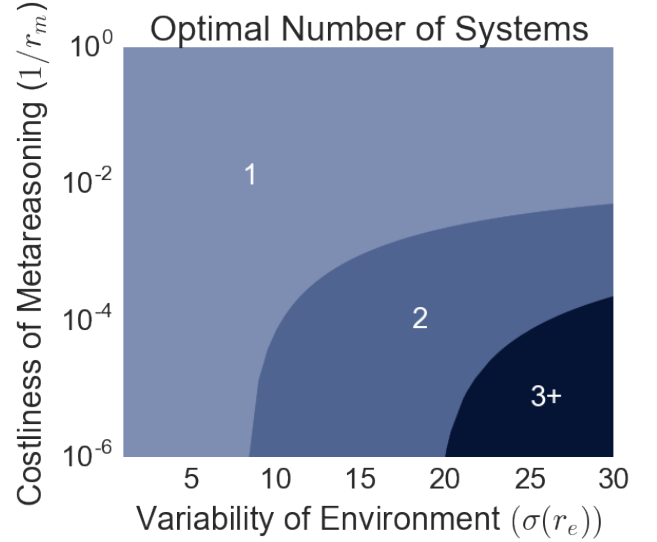

Figure 2: The optimal number of systems for the 2AFC problem as a function of the standard deviation of r_e and $1/r_m$. In this example $\mathbb{E}[r_e] = 10$.

Note that the optimal set of systems depends on the distribution of the sampling cost r_e across different environments. In the following we assume that $r_e - 1 \sim \Gamma(\alpha, \beta)$ was sampled from a Gamma distribution so that acting is always as least as costly as sampling ($r_e \geq 1$).

Figure 1 shows a representative example[1] of the expected utility per time as a function of the number of systems for different metareasoning costs. Note that under a large range of metareasoning costs the optimal number of systems is just one, but as the costliness of metareasoning 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 costliness of metareasoning (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 has high variability but there is also a correspondingly

---

[1] For all experiments reported in this paper, we found that alternative values for $\mathbb{E}[r_e]$ or $\text{Var}(r_e)$ did not change the qualitative conclusions, unless otherwise indicated.
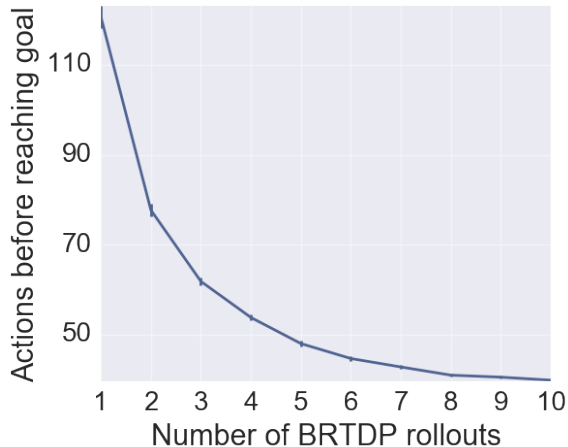
Figure 3: The number of actions it takes an agent to reach a goal as a function of the number of BRTDP rollouts executed before each action. For 0 rollouts the expected number of actions was 500 (the maximum allowed).

high enough cost of metareasoning that the optimal number of cognitive systems is only two.

Planning under Uncertainty

When an agent is unable to precompute an optimal policy for an environment it must interleave planning with acting. However, planning in the real world comes with a situation-dependent cost, so an optimal agent must metareason to adaptively choose how much to plan. The problem of choosing how to choose the optimal sequence of computations to plan which actions to take in the MDP corresponds to a meta-level MDP (Hay et al. 2012).

Here, we study an agent solving a stochastic shortest path problem in a finite-horizon MDP with states S, actions A, cost function $c : S \to \mathbb{R}$, transition probability $t : S \times A \times S \to [0, 1]$, and horizon h. In the meta-level decision process each time step consists of a thinking stage followed by an acting stage. The agent is equipped with a set of planning algorithms $\mathcal{T}$, and the cost of running an algorithm $t \in \mathcal{T}$ is given by $f(t)$. The optimal policy in this metareasoning problem can be expressed in terms of the trajectory ξ of the states and chosen planning algorithms at each time step:

$$\pi^* = \arg\max_{\pi} \mathbb{E} \left[\sum_{(s,t) \in \xi} c(s) + f(t) \Big| \pi \right], \qquad (7)$$

where ξ depends on the agent's metalevel policy, π.

Our agent's thinking algorithms are based on bounded real-time dynamic programming (BRTDP; McMahan, Likhachev, and Gordan, 2005), an anytime online planning algorithm for solving an MDP, that was previously used by Lin et al. to metareason for planning under uncertainty. BRTDP starts with an initial heuristic for a lower and upper bound on the value function of the MDP and updates its estimate of the bounds via rollouts. As more rollouts are

done, the gap between the lower and upper bound decreases, eventually converging to the optimal value function.

Because optimally solving a meta-level decision problem is generally computationally intractable (Lin et al. 2015), we focus on a special case where the agent's thinking algorithms differ only in the number of rollouts (k) they perform before each action. During the acting stage, the agent acts greedily with respect to the upper bound. Thus the agent's policy is defined entirely by k. This type of policy corresponds to the Think*Act policy from Lin et al..

We consider environments with a constant cost for each action (c_a) and assume a constant execution cost (c_e) for each rollout of BRTDP. We can therefore reparameterize the costs by the ratio of the cost of acting over the cost of thinking, $r_e = \frac{c_a}{c_e}$. This simplifies Equation 7 to

$$k^* = \arg\min_{k \in \mathbb{N}_0} \left(1 + \frac{k}{r_e} \right) \mathbb{E}\left[N|k\right], \qquad (8)$$

where N is the number of actions taken in the MDP until reaching the goal state or time horizon.

We assume that there is a distribution of potential MDPs that our agent may act in, and while r_e is constant within each MDP, it varies across different MDPs. Therefore, optimally allocating finite computational resources requires metareasoning. Metareasoning incurs a cost, c_m, that can be similarly reparameterized as $r_m = c_a/c_m$. Assuming that the agent chooses optimally from its set of planning systems, the optimal set of systems that it should be equipped with is

$$\mathcal{M}^* = \arg\min_{\mathcal{M} \subset \mathbb{N}} \mathbb{E}_{r_e} \left[\min_{k \in \mathcal{M} \cup \{0\}} \frac{(r_e + k)\mathbb{E}\left[N|k\right]}{r_e} \right] + \frac{|\mathcal{M}|}{r_m}. \quad (9)$$

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, the maximum number of rollouts at any thinking stage to 10, and the depth of each BRTDP rollout 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 (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).

We find that all our results match the 2AFC setting extremely closely. Because the agent rarely reached the goal without 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 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
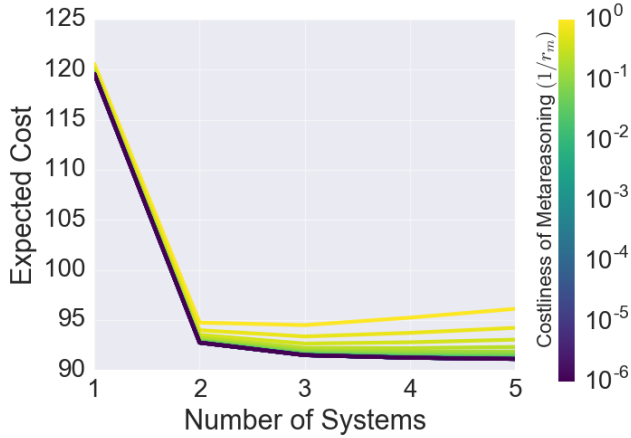
Figure 4: The expected cost incurred is a U-shaped function of the number of planning systems. As the costliness of metareasoning, $\frac{1}{r_m}$, decreases, the optimal number of systems increases. The expected cost of 0 systems was 500, thus 1 system provided the greatest reduction in cost. In this example $\mathbb{E}[r_e] = 100$, $\text{Var}(r_e) = 10^5$, and $c_a = 1$.

$\|\mathcal{M}\|$	$\text{Var}(r_e)$		
	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

Table 2: The optimal set of algorithms, $\mathcal{M}$, for different values of $\|\mathcal{M}\|$ as $\text{Var}(r_e)$ increases while $\mathbb{E}[r_e] = 100$.

increased with the variance of r_e.[2] 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.

Conclusion

We found that across two different domains the optimal number and diversity of cognitive systems increases with the variability of the environment but decreases with the cost of metareasoning. Each additional system tends to provide at most marginal improvements, so the 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 one of them was typically faster but more error-prone whereas the second one was slower but more accurate, matching the systems posited by dual-process theories.

---

[2]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.
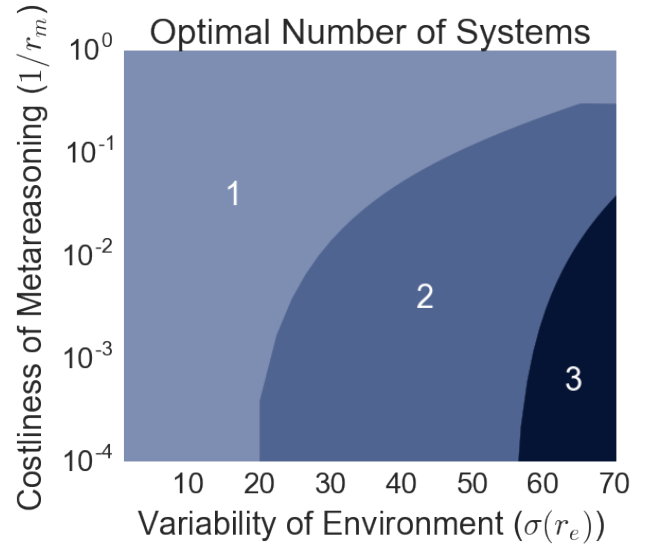

Figure 5: The optimal number planning systems as a function of the standard deviation of r_e and r_m for $\mathbb{E}[r_e] = 100$.

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 intelligent cognitive systems. A second limitation is that there is no guarantee that the results we obtained in the domain of decision-making will necessary transfer to all other domains of cognition. However, it seems plausible that for most metareasoning systems the cost of metareasoning 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 cost of metareasoning is likely to generalize to other tasks and more complex architectures.

Future work in AI might apply our approach to design bounded-optimal metareasoning agents whose performance comes closer to the impressive resource-efficiency and flexibility of the human mind. According to our analysis, bounded agents should be equipped with a small number of cognitive systems dependent on the variability agent's environment and the cost of metareasoning rather than just one or dozens of reasoning systems.

Our findings also have implications for the debate about human rationality. In this debate, some researchers have interpreted the existence of a fast, error-prone cognitive system that violated 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 additional to a slow but accurate system can be boundedly optimal.

A conclusive answer to the question whether it is boundedly optimal for humans to have two types of cognitive sys-

tems will require more rigorous estimates of the variability of decision problems that people experience in their daily lives and the cost of metareasoning. Regardless thereof, our analysis suggests that the incoherence in human reasoning and decision-making might be a signature of the rational use of a bounded-optimal set of cognitive systems rather than a sign of irrationality.

References

Anderson, M. L., and Oates, T. 2007. A review of recent research in metareasoning and metalearning. *AI Magazine* 28(1):12.

Ariely, D. 2009. *Predictably Irrational*. New York: Harper Collins.

Cox, M. T., and Raja, A. 2011. *Metareasoning: Thinking about thinking*. Cambridge, MA: MIT Press.

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

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

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

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

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., and Tenenbaum, J. B. 2006. Optimal predictions in everyday cognition. *Psychological science* 17(9):767–773.

Griffiths, T. L.; Lieder, F.; and 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.

Hay, N.; Russell, S.; Tolpin, D.; and Shimony, S. 2012. Selecting Computations: Theory and Applications. In de Freitas, N., and Murphy, K., eds., *Proceedings of the 28th conference on Uncertainty in Artificial Intelligence*. Corvallis: AUAI Press.

Horvitz, E., and Rutledge, G. 1991. Time-dependent utility and action under uncertainty. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 151–158. Morgan Kaufmann Publishers Inc.

Horvitz, E. J.; Cooper, G. F.; and 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*, 1121–1127. San Mateo, CA: Morgan Kaufmann.

Icard, T. 2014. Toward boundedly rational analysis. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 637–642.

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

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

Lieder, F., and Griffiths, T. L. 2015. When to use which heuristic: A rational solution to the strategy selection problem. In *Proceedings of the 37th annual conference of the cognitive science society*, 1362–1367.

Lieder, F.; Plunkett, D.; Hamrick, J. B.; Russell, S. J.; Hay, N.; and Griffiths, T. 2014. Algorithm selection by rational metareasoning as a model of human strategy selection. In *Advances in Neural Information Processing Systems 27*, 2870–2878.

Lieder, F.; Griffiths, T.; and Goodman, N. 2012. Burn-in, bias, and the rationality of anchoring. In *Advances in neural information processing systems 25*, 2690–2798.

Lieder, F.; Hsu, M.; and Griffiths, T. L. 2014. The high availability of extreme events serves resource-rational decision-making. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Lin, C. H.; Kolobov, A.; Kamar, E.; and Horvitz, E. 2015. Metareasoning for planning under uncertainty. In *Proceedings of the 24th International Conference on Artificial Intelligence*, 1601–1609. AAAI Press.

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

McMahan, H. B.; Likhachev, M.; and 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*, 569–576. ACM.

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

Russell, S., and 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*, 334–340. Morgan Kaufmann Publishers Inc.

Russell, S., and Wefald, E. 1991a. Principles of metareasoning. *Artificial intelligence* 49(1-3):361–395.

Russell, S. J., and Wefald, E. 1991b. *Do the right thing: studies in limited rationality*. Cambridge, MA: MIT press.

Schaul, T., and Schmidhuber, J. 2010. Metalearning. 5(6):4650. revision 91489.

Smith-Miles, K. A. 2009. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM computing surveys* 41(1).

Stanovich, K. 2011. *Rationality and the reflective mind*. Oxford: Oxford University Press.

Thornton, C.; Hutter, F.; Hoos, H. H.; and Leyton-Brown, K. 2013. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 847–855. New York: ACM.

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