Cognition as a sequential decision problem

Frederick Callaway

a dissertation
presented to the faculty
of princeton university
in candidacy for the degree
of Doctor of Philosophy

recommended for acceptance
by the Department of
Psychology

Adviser: Thomas L. Griffiths

November 2022

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Morbi commodo, ipsum sed pharetra gravida, orci magna rhoncus neque, id pulvinar odio lorem non turpis. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetuer. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros.

Quisque facilisis erat a dui. Nam malesuada ornare dolor. Cras gravida, diam sit amet rhoncus ornare, erat elit consectetuer erat, id egestas pede nibh eget odio. Proin tincidunt, velit vel porta elementum, magna diam molestie sapien, non aliquet massa pede eu diam. Aliquam iaculis. Fusce et ipsum et nulla tristique facilisis. Donec eget sem sit amet ligula viverra gravida. Etiam vehicula urna vel turpis. Suspendisse sagittis ante a urna. Morbi a est quis orci consequat rutrum. Nullam egestas feugiat felis. Integer adipiscing semper ligula. Nunc molestie, nisl sit amet cursus convallis, sapien lectus pretium metus, vitae pretium enim wisi id lectus. Donec vestibulum. Etiam vel nibh. Nulla facilisi. Mauris pharetra. Donec augue. Fusce ultrices, neque id dignissim ultrices, tellus mauris dictum elit, vel lacinia enim metus eu nunc.

Contents

This is the dedication.

Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Morbi commodo, ipsum sed pharetra gravida, orci magna rhoncus neque, id pulvinar odio lorem non turpis. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetuer. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros.

0 Introduction

How can we build theoretically satisfying and practically useful models of human cognition? Historically, there have been two broad approaches. The *rational* approach, exemplified by the work of David Marr (e.g., 1982) and John Anderson (e.g., 1990), begins by characterizing the problems people have to solve and then assumes that their cognitive systems will be optimally designed to solve those problems. Rational models are satisfying because they tell us *why* the mind works the way it does, and they are useful because they allow us to make generalizable predictions about how people will behave in new environments. However, by construction, such models don't explain *how* the mind achieves the rational ideal, and a growing list of systematic cognitive biases draw their predictive utility into question.

In contrast, the *mechanistic* approach focuses on identifying the cognitive processes underlying behavior, often with an emphasis on explaining the behavioral idiosyncrasies that rational models gloss over. This approach can potentially tell us how the mind actually works, and it can produce extremely accurate models. However, mechanistic models are often highly tuned to specific experimental setups—we are left wondering why this specific model fit data best, and whether it would continue to make good predictions in a slightly different environment.

Although the rational and mechanistic approaches have traditionally been viewed as conflicting, the past decade has seen a resurgence of an old idea (Simon, 1955): rationality can be seen as a property of cognitive mechanisms themselves. Specificially, a cognitive mechanism is rational if it makes optimal use of limited cognitive resources. Going under various names—cognitively bounded rational analysis (Howes et al., 2009), computational rationality (Lewis et al., 2014; Gershman et al., 2015), and resource-rational analysis (Griffiths et al., 2015; ?) to name a few—this view suggests that we should not expect people to be rational in the traditional sense of selecting actions that maximize utility (Von Neumann and Morgenstern, 1944). Instead, we should expect people to select actions using mental strategies that strike a good tradeoff between the utility of the chosen action and the cogntive cost of making the decision.

But what defines a "good" tradeoff between action utility and cognitive cost? And how can we identify mental strategies that achieve such a tradeoff? Here, we suggest answers to these questions based on a key insight: a rational mental strategy is one that optimally solves the sequential decision problem posed by one's internal computational environment. Under this view, cognition is a problem of stringing together a series of basic cognitive operations or "computations" in the service of choosing which actions to take in the world. By formalizing this problem as a Markov decision process (MDP, the formalism underlying reinforcement learning), we can leverage standard tools from artificial intelligence to identify optimal cognitive processes. Doing so has already revealed a number of domains in which people's cognitive strategies are remarkably close to optimal, suggesting that the approach can provide accurate and generalizable predictions about human behavior. At the same time, people often show systematic deviations from the optimal model, suggesting ways that both our models and people's mental strategies might be improved.

0.1 Optimal cognitive processes as solutions to Markov decision processes

The proposed approach rests on a key intuition: the thoughts one has at any moment depend on the thoughts one had before. That is, our mental processes are sequentially dependent. Furthermore, thoughts are only useful insofar as they influence our behavior, and this behavior often occurs well after the thought itself. That is, the benefits of thought are temporally delayed. These two properties, sequential dependence and delayed reward

make sequential decision problems very challenging to solve. Fortunately, a long history of work in artificial intelligence—from Newell and Simon's pioneering proof-writing programs(Newell and Simon, 1956) to super-human Chess and Go engines (Silver et al., 2017)—has focused on solving just this sort of problem.

In artificial intelligence research, sequential decision problems are often formalized with the framework of Markov decision processes (MDP) (Puterman, 2014; Sutton and Barto, 2018). An MDP describes a dynamic interaction between an agent and an environment. It is defined by a set of possible states the environment can be in, a set of actions the agent can execute, a reward function that specifies the immediate utility associated with executing each action in each state, and a transition function that specifies how actions change the state. The agent's goal is to maximize the total cumulative reward received. This can be accomplished by choosing actions according to the optimal policy, which specifies the best action to execute in each state. See Box 1 for a technical definition of MDPs.

The fact that cognition—or more generally, computation—poses a sequential decision problem was recognized by researchers in the field of rational metareasoning, which aims to build AI systems that can adaptively allocate their limited computational resources. In particular, Hay and colleagues Hay et al., 2012; Hay, 2016 formalize this problem of "selecting computations" as a metalevel MDP. In a metalevel MDP, the states correspond to beliefs and the actions correspond to computations that refine those beliefs (according to the transition function). The reward function encodes both the costs and benefits of computation; it assigns a strictly negative reward for each computation executed, but a potentially positive reward for the utility of the external action that is ultimately chosen (based on the belief produced by computation).

Applying the metalevel MDP formalism to cognitive science provides a suite of theoretical and computational tools, both to formalize the problems that our brains must solve, and to identify near-optimal solutions to those problems. In a psychological context, the states of a metalevel MDP can be interpreted as mental states, and the actions as cognitive operations. Along with the transition function, these specify the environment "within the skin" that a cognitive process must interact with. By further specifying a reward function, we can quantify the tradeoff between cognitive cost and extrinsic utility . This in turn allows us to identify the optimal cognitive process—that is, the one achieves the best possible tradeoff between cost and utility—as the optimal policy for the metalevel MDP.

We must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organism

Herbert Simon

1

Metalevel Markov decision processes

As outlined above, the key insight underlying our framework is that cognitive processes such as decision-making can themselves be viewed as sequential decision problems. Drawing on a subfield of artificial intelligence known as *rational metareasoning* (Matheson, 1968; Russell and Wefald, 1991), we formalize this insight using the framework of *meta-level Markov decision processes* (meta-level MDPs; Hay et al., 2012). In this framework, a cognitive process is formalized as a sequential process of executing computational actions that update an agent's beliefs about the world. At each moment, the agent must choose whether to continue deliberating, refining their beliefs but accruing computational cost, or to instead stop computing and make a decision. In the former case, they must additionally decide which computation to execute next (i.e., what to think about); in the latter case, they select the optimal action given their current belief and receive a reward associated with the external utility of that action.

In this chapter, I describe the formal framework and show how it can be applied to multi-attribute choice, the domain in which we conduct our experimental case studies. We provide a non-technical summary at the end of this section.

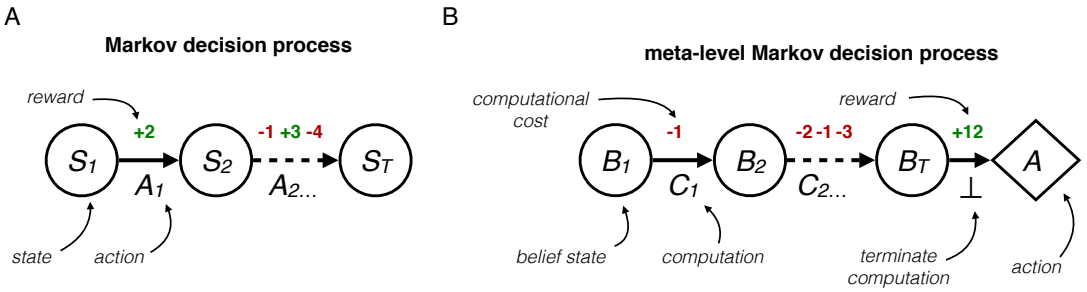

Figure 1.1: Formal framework: meta-level Markov decision processes. (A) A Markov decision process formalizes the problem of acting adaptively in a dynamic environment. The agent executes actions that change the state of the world and generate rewards, which the agent seeks to maximize. (B) A meta-level Markov decision process formalizes the problem of *deciding how to act* when computational resources are limited. The agent executes computations that update their belief state and incur computational cost. When the agent executes the termination operation $\perp$, they take an external action based on their current belief state.

1.1 Markov decision processes

The core mathematical object underlying our approach is the Markov decision process (MDP), illustrated in Figure 1.1A. MDPs are the standard formalism for modeling the sequential interaction between an agent and a stochastic environment. An MDP is defined by a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a transition function T, and a reward function r. A state $s \in \mathcal{S}$ specifies the relevant state of the world. An action $a \in \mathcal{A}$ is an action the agent can perform. The transition function T encodes the dynamics of the world as a distribution of possible future states for each possible previous state and action. Finally, the reward function r specifies the reward or utility for executing a given action in a given state.

The standard goal in an MDP is to maximize the expected cumulative reward attained, that is, the *return*. Importantly, this may require incurring immediate losses (negative rewards) in order to get to a state from which a highly rewarding action can be executed. It is typically assumed that the agent selects their actions based on the current state; the mapping from state to action is called a policy, denoted π. Solving an MDP amounts to finding a policy that maximizes the expected return, that is, a mapping from states to actions that, when followed, maximizes the total reward one will receive on average.

In addition to their foundational role in artificial intelligence (Sutton and Barto, 2018), MDPs are widely used in models of human decision-making (Dayan and Daw, 2008). MDPs are the formal foundation for models of reinforcement learning (Niv, 2009) and

model-based planning (Huys et al., 2015; Botvinick and Toussaint, 2012), as well as competition between the two systems (Daw et al., 2005; Keramati et al., 2011; Kool et al., 2017). They have also been used to study information-seeking (Gottlieb et al., 2013; Hunt et al., 2016), generalization (?), and hierarchical abstraction (Solway et al., 2014; Botvinick et al., 2009). However, with a few notable exceptions (Dayan and Huys, 2008; Drugowitsch et al., 2012; Tajima et al., 2016), MDPs have primarily been used to model the sequential decision problems posed by the external world. In the following section, we show how this framework can be applied to model the sequential decision problem posed by one's own cognitive architecture.

1.2 Meta-level Markov decision processes

Meta-level Markov decision processes (meta-level MDPs) extend the standard MDP formalism to model the sequential decision problem posed by resource-bounded computation (Hay et al., 2012). Like a standard MDP, there is a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, and a reward function r_{object} (we omit the transition function because we focus on one-shot decisions). These define the *object-level* problem: the external problem the agent must solve in the world. Additionally, the meta-level MDP defines a set of beliefs $\mathcal{B}$, a set of computations $\mathcal{C}$, and meta-level transition and reward functions, T_{meta} and r_{meta}. These define the *meta-level* problem: how to allocate limited computational resources in the service of solving the object-level problem.

As illustrated in Figure 1.1B, the meta-level problem is itself a sequential decision problem, analogous to one defined by a standard MDP. However, in the meta-level problem, the states are replaced by beliefs (mental states) and the actions are replaced by computations (cognitive operations). The meta-level transition function describes how computations update beliefs, and the meta-level reward function captures both computational cost and the object-level reward of the action that is ultimately executed. We provide a formal definition below.

We define a meta-level MDP as $(\mathcal{S}, \mathcal{A}, r_{\text{object}}, \mathcal{B}, \mathcal{C}, T_{\text{meta}}, r_{\text{meta}})$. The first three components define the task-level problem. They have the same interpretation as $\mathcal{S}$, $\mathcal{A}$ and r in a standard MDP (we omit the transition function because we limit our analysis to one-shot decisions). The latter four components define the meta-level problem. We now define these

four components in turn.

Beliefs A belief state $b \in \mathcal{B}$ captures the agent's current knowledge about the relevant state of the world. Formally, a belief is a distribution states, $\mathcal{B} \subseteq \Delta(\mathcal{S})$. Note that $\Delta(\mathcal{S})$ denotes the set of all possible distributions over $\mathcal{S}$. Importantly, contrary to a standard rational treatment of beliefs, the belief states in a meta-level MDP do not include all the information that is available to the DM. Instead, the belief state only contains information that is immediately accessible, excluding, for example, long-term memories and the number of calories in every box of cereal on a shelf.

Computations A computational operation $c \in \mathcal{C}$ is a primitive operation afforded by the computational architecture. Formally, it is a meta-level action that updates the belief in much the same way as a regular action changes state. All meta-level MDPs include the termination operation $\perp$, which denotes that computation should be terminated and an action should be selected based on the current belief state. We further explain belief updating and termination in the following two paragraphs.

Transition function The meta-level transition function $T_{\text{meta}} : \mathcal{B} \times \mathcal{C} \times \mathcal{S} \to \Delta(\mathcal{B})$ describes how computation updates beliefs. At each time step, the next belief is sampled from a distribution that depends on the current belief, the computational operation that was just executed, and the true state of the world, that is,

$$b_{t+1} \sim T_{\text{meta}}(b_t, c_t, s). \tag{1.1}$$

The transition function thus defines the core structure of the computational architecture. Following previous work (??), we assume that the effect of computation is to generate or reveal information about the true state of the world, which is then integrated into the belief state. Thus, in expectation, computation has the effect of making one's beliefs more precise and accurate, although an individual computation may yield misleading information.

Reward function The meta-level reward function $r_{\text{meta}} : \mathcal{B} \times \mathcal{C} \times \mathcal{S} \to \mathbb{R}$ describes both the costs and benefits of computation. For the former, r_{meta} assigns a strictly negative

reward for all non-terminating computational operations,

$$r_{\text{meta}}(b, c, s) = -\text{cost}(c) \text{ for } c \neq \bot. \tag{1.2}$$

The cost of computation may include multiple factors, such as energetic costs and opportunity costs.

Intuitively, the benefit of computation is that it allows one to make better decisions. This is captured by the meta-level reward for the termination operation $\bot$, defined as the true utility of the external action that the DM would execute given the current belief. We assume that the action is selected optimally. Thus,

$$r_{\text{meta}}(b, \bot, s) = r_{\text{object}}(s, a^*(b)). \tag{1.3}$$

where

$$a^*(b) = \operatorname*{argmax}_a \mathrm{E}\left[r_{\text{object}}(s, a) \,\middle|\, s \sim b\right] \tag{1.4}$$

In English, the meta-level reward for termination is the *true* utility of the action* with maximal *estimated* utility.

Policy The solution to a meta-level MDP takes the form of a policy $\pi : \mathcal{B} \to \Delta(\mathcal{C})$ that (perhaps stochastically) selects which computation to perform in each possible belief state. The optimal policy is the one that maximizes expected meta-level return,

$$\pi^* = \operatorname{argmax} \mathrm{E}\left[\sum_{t=1}^{T} r_{\text{meta}}(B_t, C_t, S) \,\middle|\, C_t \sim \pi\right]. \tag{1.5}$$

*For notational clarity, we have assumed a single optimal action. When multiple actions have the same expected value, we assume that ties are broken randomly; thus, $a^*(b)$ is more precisely a uniform distribution over all optimal actions, and $r_{\text{meta}}(b, \bot, s)$ takes an expectation over them.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Psychology Press.

Botvinick, M. and Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, 16(10):485–488.

Botvinick, M. M., Niv, Y., and Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3):262–280.

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

Dayan, P. and Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective and Behavioral Neuroscience*, 8(4):429–453.

Dayan, P. and Huys, Q. J. M. (2008). Serotonin, Inhibition, and Negative Mood. *PLOS Computational Biology*, 4(2):e4.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. *Journal of Neuroscience*, 32(11):3612–3628.

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.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11).

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. (2016). Principles of Metalevel Control.

Hay, N., Russell, S., Tolpin, D., and Shimony, S. E. (2012). Selecting computations: Theory and applications. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, UAI'12, pages 346–355, Arlington, Virginia, USA. AUAI Press.

Howes, A., Lewis, R. L., and Vera, A. (2009). Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action. *Psychological Review*, 116(4):717–751.

Hunt, L. T., Rutledge, R. B., Malalasekera, W. M. N., Kennerley, S. W., and Dolan, R. J. (2016). Approach-Induced Biases in Human Information Sampling. *PLOS Biology*, 14(11):e2000638.

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., and Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10):3098–103.

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.

Kool, W., Gershman, S. J., and Cushman, F. A. (2017). Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychological Science*, 28(9):1321–1333.

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

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: WH Freeman.

Matheson, J. E. (1968). The Economic Value of Analysis and Computation. *IEEE Transactions on Systems Science and Cybernetics*, 4(3):325–332.

Newell, A. and Simon, H. (1956). The logic theory machine–A complex information processing system. *IRE Transactions on Information Theory*, 2(3):61–79.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

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

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van Den Driessche, G., Graepel, T., and Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., and Botvinick, M. M. (2014). Optimal Behavioral Hierarchy. *PLOS Computational Biology*, 10(8):e1003779.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT press.

Tajima, S., Drugowitsch, J., and Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, 7(1):12400.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, US.