The Cost of Decision-Making:
Unraveling the Interplay of Memory, Reinforcement Learning, and Neural Connectivity in Decision-Making

Yuxue Yang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:
Andrea Stocco, Chair
Chantel Prat
Brian Flaherty
Catherine Sibert

Program Authorized to Offer Degree:
Psychology

University of Washington

Abstract

The Cost of Decision-Making: Unraveling the Interplay of Memory, Reinforcement Learning, and Neural Connectivity in Decision-Making

Yuxue Yang

Chair of the Supervisory Committee:
Andrea Stocco
Department of Psychology

This dissertation reviewed significant theoretical work and empirical findings in the field of decision-making, and posed critical questions spanning various aspects of decision-making, such as cognitive control, motivation, individual differences, memory, and computational modeling. To address these questions, I present three computational works, aiming to improve understanding of decision-making processes within the ACT-R cognitive framework. The first computational work outlines a novel framework for cognitive effort allocation, proposing that optimal cognitive control is based on a balance of cognitive effort and benefits. The second piece of research identifies two decision-making strategies: memory-based and reinforcement-based, and uncovers how individuals' preferences for one strategy over another are predicted by their resting-state brain connectivity. By employing computational cognitive models and ML models, this study demonstrated that stronger connectivity in frontoparietal and memory retrieval regions corresponds to a preference for memory-based strategies, while stronger connectivity in sensorimotor, cingulate, and basal ganglia regions predicts a preference for reinforcement-based strategies. This work highlights the adaptive nature of human decision-making in both behavioral and neurofunctional data. Lastly, the third piece of computational research reconciles two primary learning approaches in complex decision-making tasks, introducing a cognitively plausible model that combines memory retrieval and model-based reinforcement learning (MB-RL). This integrated model connects established findings in reinforcement learning and offers insights into the balance between memory and reinforcement learning in decision-making, paving the way for advancements in artificial intelligence and decision-making modeling.

TABLE OF CONTENTS

LIST OF FIGURES1

LIST OF TABLES

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Andrea Stocco, for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me throughout the time of research and writing of this thesis. I would also like to thank the rest of my thesis committee: Dr. Chantel Prat, Dr. Catherine Sibert, for their insightful comments and encouragement, but also for the hard questions which incited me to widen my research from various perspectives. My sincere thanks also go to the University of Washington, Psychology Department, Cognition & Cortical Dynamic Laboratory for providing me with an excellent atmosphere for doing research. I thank my fellow researchers and friends in the UW for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years.

I would like to express my profound gratitude to my mother, Shuying Xiang, an exceptional woman who embodies strength and independence, serving both as a physician and a single mom, always encouraged me to pursue any career I was passionate about, providing unwavering emotional support at all times; and my beloved pets, Kuchacha and Mona, whose comforting presence helped soothe many of my unsettled and stressful nights. Last but not least, a special thanks to Zhenda Lu, my significant other, and my companion in intellectual pursuits. His profound knowledge and thought-provoking conversations have been a source of continual learning. As a boyfriend and a dear friend, his unconditional love and support have been invaluable throughout my five-year journey in graduate school.

DEDICATION

To my mentor, my family, and my furry friends.

I couldn’t have done this without your unwavering support and love.

Chapter 1. INTRODUCTION

1.1 OVERVIEW OF DISSERTATION

The field of decision-making has made significant progress in recent years, but several unsolved questions remain. These questions span different aspects of decision-making, including cognitive control, motivation, individual differences, memory, as well as computational modeling. It is necessary to advance the understanding of the cognitive foundations of decision-making, especially from a cognitive modeling perspective. While a large body of research has delved into the observable and neurological aspects of decision-making, a comprehensive understanding linking these cognitive components remains incomplete. I am particularly interested in how cognitive costs, motivation, and memory shape our decision-making strategies and how factors at the individual and situational level influence these processes. For example, how do personal factors like cognitive costs, neural flexibility, and memory characteristics interplay with contextual factors like task difficulty or environmental uncertainty? By navigating these complex questions, we aim to construct a more complete depiction of human decision-making. We also need to discern how individual-level factors (e.g. cognitive costs, neural activity, and memory characteristic etc.) and context-specific factors (e.g. task difficulty, uncertainty of the environment etc.) modulate these processes respectively. Only by addressing these intricate questions can we hope to develop a more comprehensive and realistic model of human decision-making.

Based on this, my dissertation thesis attempts to unpack the above core questions and focus on modeling the cognitive processes underlying decision-making within a cognitive architecture. It is commonly believed that decision-making is a cognitively demanding process, and understanding how cognitive effort is allocated to make optimal decisions is of particular interest.

This dissertation provides the first attempt to answer how the internal costs associated with cognitive computations affect decision-making. Specifically, in Chapter 1, I begin with an introduction to key cognitive processes in decision-making and explore the role of cognitive effort in decision-making, as well as the neural correlates of cognitive effort and decision-making. Then I will introduce commonly used measures of cognitive effort, further discuss how to study individual differences in cognitive effort, and review prominent theories of optimal decision-making.

In Chapter 2, I will delve into prominent computational frameworks that are important in modeling decision-making and effort allocation, such as reinforcement learning, the Expected Value of Control (EVC) theory, and ACT-R architecture. Following this, three empirical studies are presented in Chapter 3-5 to answer a series of questions:

1. How does the Expected Value of Control (EVC) theory explain the selection of cognitive processes an individual would engage when carrying out a cognitive task?
2. Can we predict whether different people would carry out the same task in different ways, based on the individual-level efficiency of the corresponding brain circuits?
3. What combination of cognitive processes will be used when individuals carry out *complex* decision-making tasks, such as those that require planning ahead?

To address these questions, in the first study, I will explore the nature of mental rewards and costs and investigate the mechanisms of cognitive resource allocation in a classic paradigm, the Simon Task (1955). In the second study of my thesis, I will examine whether people rely on the same or different cognitive processes to make decisions under such conditions. I will investigate the possibility of predicting individual differences in strategy use from neural connectivity, aiming to understand the adaptive nature of decision-making and the factors driving these differences. In the third study, I will delve into the role of memory retrieval in the “planning” aspect of decision-making. As the degree of planning increases with age, my model accounts for developmental changes of planning within a cognitive architecture. I will explore whether memory decay can capture the developmental change in more effortful decision-making strategies.

In the final chapter, I will tie together the results of these studies and propose a united framework for how the costs associated with different cognitive processes shape individual differences in decision making. I’ll conclude with a comprehensive discourse on the significance, limitations, and future direction in this field, thus contributing valuable insights into our understanding of decision-making processes and the elements that shape them.

1.2 System 1 and System 2 in Decision-Making

Decision making is not a single, monolithic process, and several alternative mechanisms for decision-making have been proposed. Perhaps the most famous is the distinction between “System 1” and “System 2”, popularized by Kahneman (2011), in *"Thinking, Fast and Slow,"* and originally proposed by Evans in 2008. *System 1* represents fast, intuitive, and automatic thinking, while *System 2* corresponds to slower, more deliberative, and effortful thinking.

System 1 and 2 can be connected to the procedural and declarative cognitive frameworks respectively (Singley & Anderson, 1989). In cognitive literature, the declarative and procedural frameworks are two key ways we learn and remember things (Evans, J. St. B. T., & Stanovich, K. E. (2013). Declarative learning involves conscious recall of facts and experiences, like remembering a movie plot. Procedural learning, on the other hand, is about gaining skills through practice, like learning to ride a bike(Squire, 2004). While it’s common to correlate System 1 and System 2 with procedural and declarative knowledge respectively, not everyone is in agreement with this view. A recent preprint by Conway-Smith and West (2023) has even proposed a different interpretation. They propose a revised understanding of System 1 and System 2 within the framework of the Common Model of Cognition, arguing that the two systems aren’t exclusively tied to the procedural and declarative systems as traditionally assumed. Their work emphasizes the shared characteristics of working memory linked with System 1 and System 2, and particularly pointed out that even if a System 2 process is predominantly guided by declarative knowledge, it would still require the procedural knowledge from System 1 for the retrieval and execution of actions. This perspective further highlights the complexity of our cognitive systems and indicates that our understanding of decision-making processes is still evolving.

These two ways of learning shape much of our understanding of cognition and decision-making. Just as procedural memory governs our ability to perform certain tasks without conscious thought, *System 1* handles automatic cognitive tasks, such as identifying objects or recognizing facial expressions. These processes are typically fast, efficient, and require minimal mental effort. On the other hand, *System 2* thinking, characterized by deliberate, analytical, and slow cognitive processes, can be likened to the declarative framework, akin to the conscious and effortful problem-solving. This kind of cognition involves evaluating information, reasoning, and making deliberate decisions, which are more time-consuming and mentally taxing than the automatic processes handled by System 1 or the procedural framework. In this context, understanding the interplay between these two systems or frameworks can provide insight into the complexity and versatility of human cognition, enhancing our understanding of decision-making, learning, and memory. Kahneman's exploration of these dual systems provides valuable insights into the cognitive processes that underlie decision-making and the factors that influence their effectiveness and efficiency.

The distinction between System 1 and System 2 thinking is applicable in every aspect of life. Imagine a college student named Mona, who is tasked with completing two assignments for her courses: a multiple-choice quiz for her introductory psychology class and a comprehensive research paper for her advanced economics course. The multiple-choice quiz requires relatively little effort, as Mona simply needs to look-up the psychology terms and recognize the correct answers among the given options. This task provides her with cognitive ease, as she can rely on her memory and intuition to make her selections. Mona feels confident and comfortable while completing the quiz, experiencing a sense of fluency and satisfaction.

On the other hand, writing a research essay demands significant cognitive effort, as Mona must conduct in-depth research, analyze complex data, and synthesize her findings into a coherent and well-structured argument. This task induces cognitive strain, as Mona has to engage in critical thinking, planning, and problem-solving to effectively complete the assignment. She feels challenged and may even experience some frustration or anxiety as she works through the various stages of the paper.

In this scenario, Mona might be tempted to procrastinate on her research paper and instead focus on tasks that offer cognitive ease, such as the multiple-choice quiz or other less demanding assignments. People have a natural tendency to gravitate toward activities that take less mental work, which explains why this behavior is so prevalent in our modern society: we want to avoid mental strain at high costs. However, to succeed in college and develop her intellectual abilities, Mona must learn to manage her cognitive resources effectively and find a balance between tasks that offer cognitive ease and those that demand cognitive strain.

The concepts of cognitive ease and cognitive strain, which are associated with cognitive endeavor, are thoroughly reviewed in the literature on cognitive effort by Kool and Botvinick (2018). When faced with a task that requires little cognitive effort, individuals experience cognitive ease. In contrast, cognitive strain occurs when a task demands significant cognitive effort. People often avoid cognitive strain and prefer tasks that demand less cognitive effort. Kool and Botvinick (2014) discuss the role of cognitive control in the allocation of cognitive effort. Cognitive control is the set of processes that facilitate flexible, goal-directed behavior, and is believed to be resource-intensive, thereby contributing to the subjective experience of cognitive effort. They further point out that although the preference for cognitive ease can lead to suboptimal decision-making, it may also have adaptive benefits, as it allows individuals to conserve cognitive resources and maintain a balance between effort expenditure and cognitive capacity.

So, when we think about *System 1* and *System 2* in terms of mental effort, we can say that System 2 tasks are generally more effortful than those handled by *System 1*. Because firstly, it often involves multiple steps or layers of thinking that need to be processed sequentially, which requires sustained focus. Secondly, it requires us to inhibit our intuitive or automatic responses (which are driven by *System 1*) and instead rely on more deliberative and thoughtful processes. Lastly, *System 2* usually is dealing with a higher degree of uncertainty or ambiguity, which makes them cognitively challenging. On the other hand, *System 1* operations tend to be more automatic and effortless. It involves routine, everyday decisions and actions, like driving on a familiar route or making quick, gut decisions based on immediate perception. This system is less taxing on our cognitive resources, as it relies on heuristics, habits, and pre-existing schemas or knowledge.

The mental effort that *System 2* requires has a biological cost—mental effort often translates to physical depletion, metabolic fatigue, and opportunity costs in multiple options because thinking hard can use up energy resources. But it's also important to note that human beings are goal-oriented animals. This exertion of mental effort has its payoffs, in that it can lead to better, more informed outcomes from the environment. In Chapter 3, I will delve deeper into this dichotomy, exploring how these two decision-making systems are distinct in their neurological and behavioral patterns. In Chapter 4-5, I will provide a more nuanced understanding of the interplay between effortful and effortless decision-making processes, and how they shape our actions and experiences.

1.3 Key Cognitive Processes in Decision-Making

As the distinction between System 1 and System 2 has shown, decision-making is a complex cognitive process that involves several key components. From a cognitive psychology perspective, one makes even finer-grained distinctions. First, the process typically begins with the identification and representation of the problem or choice at hand. This involves recognizing the available options and their potential consequences. Second, individuals must gather and evaluate relevant information about each option, weighing the potential risks and rewards. This information evaluation step often requires individuals to rely on their memory and prior knowledge, as well as their ability to search for and process new information. Third, individuals must integrate this information and formulate a preference or judgment, which involves the comparison and ranking of options based on their perceived value. Finally, individuals need to select and execute the chosen option, which may involve further planning and goal-directed action.

Thinking about a scenario of Mona, who has an important exam in the afternoon and is feeling a bit exhausted. She is considering whether to have a cup of coffee to help her stay alert during the exam. First, she identifies the goal: she needs to stay awake and focused during her exam. She recognizes that having a cup of coffee is one available option, while other options might include taking a power nap or doing a quick workout to boost her energy levels. Next, Mona gathers and evaluates relevant information about each option. She recalls from her memory that coffee has helped her stay alert in the past, but she also considers the potential risks, such as increased anxiety or jitters during exams. Additionally, she searches for information on the potential benefits and costs of a power nap or a quick workout. After evaluating the possible choices, Mona integrates it and formulates a preference. She compares the options based on their perceived value in helping her stay awake and focused during the exam. She may weigh the potential negative side effects of each choice against the potential benefits. Finally, Mona selects and executes her chosen option. Based on her analysis, she decides to have a cup of coffee, believing that it will provide her with the energy boost she needs without consuming too much time. She then proceeds to make the coffee and enjoys it before her exam. Throughout this decision-making process, Mona engages cognitive processes including attention, calculation, memory, and reasoning to make the most optimal choice for her situation.

1.4 Cognitive Effort and Decision-Making

Definition of Cognitive Effort. Cognitive effort, a critical aspect of the cognitive process involved in decision-making, refers to the amount of cognitive resources, such as attention and working memory, that an individual allocates to perform a particular task or make a decision

(Kool & Botvinick, 2018). Decision-making is a complex cognitive process that requires individuals to evaluate multiple options, weigh their respective pros and cons, and ultimately select the most beneficial course of action. Cognitive effort plays a significant role in this process, as the cognitive resources required for making a decision may vary depending on factors such as the complexity of the task, the individual's cognitive abilities, and the context in which the decision is made.

Importance of Studying Cognitive Effort. Understanding the role of cognitive effort in decision-making processes is crucial for several reasons. First, it enables researchers to identify the factors that influence the allocation of cognitive resources during decision-making and develop methods to optimize cognitive effort. This understanding can help to improve decision-making outcomes across various domains, such as education, healthcare, business, and public policy. Second, insights into the role of cognitive effort can inform the development of more accurate and cognitively plausible models of human decision-making, which are essential for advancing our knowledge of cognitive mechanisms and designing artificial intelligence systems that can mimic human-like decision-making processes. Lastly, understanding how cognitive effort shapes decision-making can help individuals become more aware of their cognitive limitations, potentially leading to better self-regulation and more efficient decision-making strategies.

Factors Impacting Cognitive Effort. Cognitive effort reflects the amount of cognitive resources allocated to perform these processes. When a decision-making task demands a high level of cognitive effort, it suggests that the cognitive processes required to make the decision are complex and resource-intensive. For instance, evaluating multiple options with numerous attributes, or dealing with ambiguous or uncertain information, can increase the cognitive effort required for decision-making. Furthermore, cognitive processes such as working memory, attention, and inhibitory control may be taxed when individuals need to keep relevant information in mind, focus on relevant cues while ignoring irrelevant information, or suppress impulsive responses in favor of more deliberate and rational choices. As a result, the relationship between cognitive effort and decision-making highlights the importance of understanding the cognitive demands associated with various decision-making tasks and the individual's capacity to allocate cognitive resources effectively to make optimal decisions.

Individual factors can significantly influence the cognitive effort required for decision-making. Cognitive abilities, such as working memory capacity, attention, and intelligence, play a vital role in determining how much cognitive effort an individual needs to invest when making decisions. Individuals with higher cognitive abilities may find it easier to process complex information and make decisions more efficiently, whereas those with lower cognitive abilities may struggle and require more cognitive effort. Personality traits, such as conscientiousness and openness to experience, can also impact decision-making processes. Highly conscientious individuals may be more prone to thoroughly evaluate their options, whereas those with high openness to experience may be more likely to consider novel or unconventional alternatives. Moreover, an individual's motivation can greatly affect the cognitive effort they are willing to invest in decision-making. Intrinsic motivation, or the drive to engage in an activity for its own sake, can lead to increased cognitive effort, while extrinsic motivation, or the drive to engage in an activity for external rewards, may yield variable results.

Task-related factors can also contribute to the cognitive effort needed for decision-making. Task complexity, which refers to the number of elements, their interrelations, and the difficulty of processing the information required to solve a problem or make a decision, can directly impact cognitive effort. As the complexity of a task increases, so does the cognitive load, leading to higher cognitive effort. Time pressure can further exacerbate this effect, as individuals are forced to make decisions under constraints, often resulting in increased cognitive effort or the adoption of simplifying strategies to reduce cognitive load. Additionally, uncertainty in decision-making situations, such as incomplete or ambiguous information, can also increase the cognitive effort required to evaluate potential outcomes and make decisions.

Contextual factors, including environmental, social, and cultural influences, can also shape the cognitive effort involved in decision-making. Environmental factors, such as noise, temperature, and lighting, can affect an individual's cognitive performance and thus influence the cognitive effort required for decision-making. Social factors, such as peer pressure, social norms, and group dynamics, can also impact decision-making processes by influencing the cognitive resources and strategies an individual employs. Cultural factors, including values, beliefs, and customs, can shape decision-making preferences and contribute to variations in cognitive effort across different cultural groups. Understanding these contextual factors is essential for developing a comprehensive understanding of the interplay between cognitive effort and decision-making.

1.5 Neural Correlates of Cognitive Effort and Value-Based Decision-Making

As the previous sections have shown, the interactions between cognitive effort and decision-making are multiple and multi-faceted. The tools of modern neuroscience, and especially the use of neuroimaging, can be used to further illuminate the interactions between these two concepts—and the regions of the brain where they interface with each other. Four regions in particular: the anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex

(DLPFC), the orbitofrontal cortex (OFC) and the striatum, are consistently highlighted in the literature as playing a role in allocation of cognitive effort in decision-making. Together, these insights form a comprehensive view that enables more accurate and robust models of decision-making, thereby enriching our contributions to cognitive science and artificial intelligence. In the next section, I will delve into a detailed review of the crucial neurobiological findings regarding the role of these brain regions in cognitive control, effort allocation, and other fundamental components of decision-making.

1.5.1 *Anterior Cingulate Cortex (ACC)*

The anterior cingulate cortex (ACC), located at the inner surface of the frontal lobe, plays a pivotal role in a range of cognitive functions, including cognitive control, attention allocation, action monitoring, error detection, inhibition, planning for higher-level goals, and emotional regulation (Botvinick and Braver, 2015). Research has suggested that the dorsal ACC is primarily responsible for executive control-related functions, while the ventral ACC is involved in emotion regulation (Bush et al., 2000).

Moreover, Kennerley et al. (2011) found that the ACC neurons encoded both the expected value and the risk associated with potential rewards, while neurons in the OFC encoded relative values of chosen choice to the recent history of choice. Similarly, Behrens et al. (2007) showed that the ACC tracks uncertainty related to decision-making, enabling the adjustment of learning rates in response to changing environments. In addition, Holroyd and Yeung (2012) proposed that the ACC plays a central role in learning from errors and reinforcing correct actions.

The ACC has been widely recognized for its role in reward-based decision-making and motivation. Kool et al. (2010) conducted a study to explore the role of the ACC in cognitive demand sensitivity and the allocation of cognitive resources to effortful tasks. Their neuroimaging results revealed that ACC activity increased as a function of cognitive demand, suggesting that the ACC plays a crucial role in monitoring and processing information related to the effort required for a task. Moreover, they found that ACC activity was modulated by individual differences in cognitive capacity, with greater ACC activation observed in participants who had higher working memory capacities.

In this context, Shenhav et al. (2013) proposed the Expected Value of Control (EVC) model, which identifies the dorsal ACC as the neural basis for calculating expected net value of exerting cognitive control for a specific action by taking into account expected reward, efficacy, and cognitive costs. This computational model has successfully explained the interaction between motivation and cognitive control, even in cases of depression (Grahek et al., 2019). Grahek et al., revealed that individuals with major depressive disorder exhibited reduced EVC signals in the dorsal ACC compared to healthy controls, which was associated with diminished cognitive control performance. Moreover, the severity of depressive symptoms was negatively correlated with the strength of the EVC signal in the dorsal ACC. These results suggest that depression might impair the ability of the dorsal ACC to accurately calculate the expected value of control, leading to suboptimal cognitive control and motivation.

Taken together, the findings of Kool et al. (2010) provide compelling evidence for the role of the ACC in cognitive demand sensitivity and the allocation of cognitive resources to effortful tasks. Neuroimaging research has provided substantial evidence supporting the central role of the ACC in monitoring cognitive demand and optimizing resource allocation, contributing to our understanding of the neural mechanisms underlying cognitive control, cognitive effort, and decision-making.

1.5.2 *Dorsolateral Prefrontal Cortex (DLPFC)*

The dorsolateral prefrontal cortex (DLPFC), located bilaterally in the upper and outer parts of the frontal lobes, has been implicated in various cognitive processes, such as working memory, cognitive control, and decision-making. Recent neuroimaging studies have expanded our understanding of the DLPFC's involvement in reward learning and cognitive effort allocation. It is believed that the ACC, the orbitofrontal cortex (OFC, located on the floor of the frontal lobes of the brain, just above the eyes), and the DLPFC together monitor and adjust the level of control by evaluating the expected reward (Braver et al.,2014). Similar to the ACC, lateral PFC is also responsible for integrating rewards and costs into decision-making analysis. The DLPFC is responsible for evaluating the reward for each action and making a decision based on all information and computations. The DLPFC has long been seen involving working memory capacity, response selection, intelligence.

Botvinick et al. (2004) conducted a study to investigate the interaction between the DLPFC and the ACC during effortful cognitive processing. Their results indicated that the DLPFC and ACC exhibited increased functional connectivity when participants performed tasks that required cognitive effort, suggesting that these two regions work together to allocate cognitive resources and optimize decision-making. Nee and D'Esposito (2016) investigated the hierarchical organization of the lateral prefrontal cortex (including the DLPFC) and its role in cognitive control. They found that the ACC and the DLPFC exhibit differential roles in conflict monitoring and cognitive control, with the ACC being more involved in detecting conflicts and the DLPFC playing a role in resolving them. This research provides further evidence for the involvement of both the ACC and the DLPFC in cognitive control processes.

In a study by Jimura et al. (2010), the authors investigated the role of the DLPFC in value-based decision-making. Participants were asked to perform a task that involved making decisions between two alternatives associated with different reward magnitudes and probabilities. The fMRI results showed that the DLPFC was engaged during decision-making and that its activity correlated with the subjective values of the chosen options. This study highlights the involvement of the DLPFC in integrating reward-related information for optimal decision-making.

A study by Westbrook et al. (2013) further examined the DLPFC's role in cognitive effort allocation. The authors used fMRI to investigate how the DLPFC processes the trade-off between cognitive effort and reward. Participants performed a task that required them to choose between high-effort, high-reward options and low-effort, low-reward options. Results suggested that the DLPFC encoded the subjective value of effort, with its activity modulated by individual differences in cognitive capacity and preferences for effortful tasks. This study demonstrates the involvement of the DLPFC in balancing the costs and benefits of cognitive effort during decision-making.

In sum, the DLPFC has been implicated in various cognitive processes, such as working memory, cognitive control, and decision-making. Recent advancements in neuroimaging have significantly broadened our comprehension of how it contributes to reward learning and the allocation of cognitive effort.

1.5.3 *Orbitofrontal Cortex (OFC)*

After receiving, processing, and encoding rewards, the brain needs to make decisions based on the collected information. The Orbitofrontal Cortex (OFC), located in the frontal lobes, just above the orbits (eye sockets), is responsible for making value-based decisions. Especially, the medial OFC is sensitive to a positive outcome, reward, while the lateral OFC responds to a negative outcome, punishment. More importantly, one study found that the OFC responds to the higher relative reward rather than the absolute reward (Tremblay and Schultz, 1999), suggesting that reward is subjective and relative in human cognition. That is to say, no single standard value of a decision could be made universally without considering one's prior experience and specific circumstances.

Moreover, the OFC, amygdala, and NAcc are three central brain regions involved in both primary (food, sex, and thirst) and secondary (money, social praise, and achievement) reward processing. NAcc, as a pleasure center, is activated when various types of rewards are shown or received. Amygdala, on the contrary, responds to fear and negative stimuli. Converging neuroscientific findings suggest that the primary function of the OFC is integrating information, computing the expected reward value for each choice, and maintaining it in working memory (Montaggue and Berns, 2002).

Damage to the OFC is associated with difficulty in emotional regulation (Phineas Gage case), inappropriate social interaction, and high-risk-seeking behaviors in gambling (Bachara et al., 1994). Although these deficits seem not to be closely related to one's motivational network, researchers have found that when the OFC is damaged, the reward calculation and evaluation no longer work, leading one to choose impulsive and high-risk actions. Such clinical cases imply that something wrong with the reward calculation may profoundly affect many aspects of cognitive functioning.

1.5.4 *Striatum*

The striatum, located in the subcortical basal ganglia of the brain, has long been regarded as a critical neural circuit involving reward and movement control. An increasing number of neuroscientific findings revealed that dopaminergic pathways in the striatum play a significant role in modulating and gating information between motivation, cognition, and decision-making (Braver et al., 2014). Dopamine is one of the most vital neural transmitters that travels around the body and communicates essential information about cognitive control, expectation, reward learning, and voluntary motor movement. Dopamine pathways play crucial roles in reward-driven cognitive processes. Specifically, there are two main types of dopamine pathways: mesolimbic dopamine pathway, and mesocortical dopamine pathway. In the mesolimbic dopamine pathway, the reward neurotransmitter, dopamine, is produced in the ventral tegmental area (VTA), travels through the globus pallidum, and is projected to the nucleus accumbens (NAcc) in the striatum, septum, amygdala, and hippocampus. This pathway enables reward anticipation and reward learning.

Dopamine also links the medial prefrontal cortex (mPFC), the ACC, and other perirhinal cortex in the mesocortical pathway. This mesocortical pathway is responsible for reward encoding, planning, and updating the expected reward of goal-directed behavior. As for reward learning, the ventral striatum is responsible for forming an associative relationship between behavior and rewarding outcome.

The nucleus accumbens (NAcc) is located at the ventral striatum, within the basal ganglia. A crucial function of NAcc in reward processing is the pleasure center. Olds and Milner (1954) conducted a famous rat experiment showing that rats would not stop pressing the lever to stimulate NAcc until they were entirely exhausted. Without a doubt, in this experiment, rats were highly motivated to press the bar. This raises an important question: whether rats press the bar because they “want” to, or because they “like” it? Subsequent studies revealed that “liking” and “wanting” are double-disassociated functions in NAcc. In other words, the ability to experience pleasure (“liking”) and the motivation to seek rewards (“wanting”) are governed by distinct neural processes within the NAcc. More importantly, dopamine, as a critical motivation-related neurotransmitter, only increases the “wanting” of the stimulus, but does not affect the pleasure aspect, “liking” (Berridge and Valenstein, 1991; Robinson & Berridge, 2003). These results support the idea that emotion and motivation recruit separate mechanisms in the brain. A state of “liking” may only be formalized as an emotional status instead of a motivational status.

Another line of research found that NAcc involves cognitive effort related functions (Salamon et al., 2005, 2007; reviewed by Braver et al., 2014). Rat experiments demonstrated that the dopamine shortage in NAcc led to slower responses and worse performance, especially in highly effortful tasks. Moreover, a preference for the low-effort task was observed in dopamine-depleted rats. Additional evidence for NAcc’s role in cognitive effort calculation revealed that the response of NAcc was discounted as a function of cognitive effort exerted (Botvinick et al., 2009). These findings supported the role of NAcc in cognitive effort regulation, which is a critical function in the cost-benefit analysis of motivation. In Shenhav (2017)’s EVC model, cognitive effort, expected reward, and efficacy are calculated to form an expected value of control (EVC), determining how much control one is willing to allocate. The cognitive effort calculation in NAcc provides invaluable insights in connecting neuroscientific evidence to the computational model of motivation proposed by Shenhav et al.

To sum up, the ventral striatum is recruited to learn the reward-action association. The striatum and the OFC were involved in reward evaluation to maintain motivation. The value-based decision-making processes occur here. The ACCand the DLPFC are the central regions responsible for regulating motivation in cognitive control to initiate a goal-directed action. Lastly, positive and negative incentives altered the functional connectivity by increasing global efficiency and decreasing decomposability.

Having explored the roles of various brain regions in regulating control and effort calculation, the next chapter will delve into major measures of cognitive effort, which are vital for studying decision-making processes. These measures include self-report ratings of mental effort, behavioral indicators like reaction time and accuracy, and physiological assessments such as eye-tracking, EEG, and fMRI. By quantitatively assessing mental exertion and resource allocation during tasks, these measures provide valuable insights into the brain's management of cognitive demands, shedding light on the underlying neural mechanisms involved in motivation, goal-directed actions, and decision-making.

1.6 Measures of Cognitive Effort

Subjective measures. Several measurements have been used to assess cognitive effort, which can be categorized into subjective, behavioral, and physiological measures. Subjective measures involve self-report scales, such as the NASA Task Load Index (Hart & Staveland, 1988), which assesses an individual's perceived cognitive effort and workload. Another example is the Rating Scale Mental Effort (RSME; Zijlstra, 1993), which allows participants to rate their perceived effort on a single-item scale. Esposito et al. (2014) conducted a study to examine the effect of mental fatigue. They asked participants to rate the degree of exhaustion on Visual Analogue Scales (VAS), ranging from 0 - 150, before and after the exhaustion-induced 2 hours flight simulation task. The Rating Scale Mental Effort (RSME) was administered during the experiment to measure participants' subjective investment of cognitive effort. Similarly, Gergelyfi et al. (2015) adopted a Multidimensional Fatigue Inventory (MFI; developed by Gentile et al., 2003) and a Post-experimental Intrinsic Motivation Inventory (IMI) to assess the level of fatigue and intrinsic motivational status before and after the experiment. These scales used a likert-type ordinal variable and summed up to a numerical score or a range, which could be fit in the subsequent linear regression model.

While subjective measurements can provide valuable insights into the experience of cognitive effort, there are several potential limitations associated with their use. First, self-report measures are subject to biases, such as social desirability and response biases, which can influence the accuracy and reliability of the reported cognitive effort. Participants may under- or overestimate their effort levels based on what they believe is expected or acceptable, rather than providing an accurate account of their true experience. Second, the introspective nature of self-report measures assumes that individuals have accurate access to their own cognitive processes, which may not always be the case. Cognitive effort can be influenced by a variety of factors, some of which may not be consciously accessible to the individual. Third, subjective measures are often influenced by contextual factors, such as mood, motivation, and individual differences, which can introduce noise and variability into the measurement of cognitive effort. Finally, subjective measurements of cognitive effort may not always align with objective measures, such as task performance or neural correlates of effort, leading to potential discrepancies between self-reported effort and actual effort exerted. As a result, it is important for researchers to carefully consider these limitations and employ a combination of subjective and objective measures to gain a more comprehensive understanding of cognitive effort in decision-making.

Behavioral measures. Behavioral measures include response times, error rates, and task performance, as longer response times, higher error rates, and lower performance may indicate increased cognitive effort (Botvinick & Braver 2015; Botvinick & Kool 2018; Braver et al., 2014).

Physiological measures. Physiological measures encompass monitoring changes in physiological parameters, such as heart rate variability (Mulder, 1992), pupillary dilation (Beatty, 1982), and electroencephalography (EEG) activity (Gevins & Smith, 2003). Another objective way to measure motivation is physiological measurements, such as cardiovascular activity, pupil diameter measurements, electrodermal activity. Gergelyfi et al. (2015) used various physiological measurements to investigate whether the cause of mental fatigue was due to loss of motivation or due to depletion of cognitive resources. Specifically, they measured EEG signals, pupil diameter, eye blink rate, skin conductance response (SCR), as well as heart rate variability (HRV) as physiological and neurological markers of motivation.

Pupil diameter has been increasingly used as an objective measure of cognitive effort in various studies (van der Wel & van Steenbergen, 2018). This physiological measure, known as pupillometry, is based on the observation that our pupils dilate when we engage in cognitively demanding tasks. By tracking changes in pupil size, we can infer the level of cognitive effort an individual is exerting at any given moment. This can be particularly useful in studies exploring decision-making, attention, memory, and other cognitive processes. Additionally, the skin conductance response (SCR) has been used in many studies to measure the level of arousal in response to a stimulus. Gergelyfi et al. (2015) applied high and low-pass filters on SCR signals and calculated the mean, standard deviation of the amplitude according to different conditions (reward vs. no-reward). These measures can provide objective, real-time indicators of cognitive resource allocation and cognitive effort during task performance.

Another important neuroimaging measurement is functional magnetic resonance imaging (fMRI) signals. fMRI measures the local concentration of oxygenated blood in the neural tissue, formally known as the Blood Oxygen Level Dependent (BOLD) signal. Because neurons consume oxygen when firing, local capillaries respond by dilating and increasing the total amount of oxygen present, thus increasing the BOLD signal.

1.7 Strategy Selection in Decision-Making

Optimal strategy selection in decision-making refers to the process of choosing the most effective and efficient approach to solve a problem or make a decision. This concept is rooted in the understanding that individuals have access to multiple strategies or cognitive processes for decision-making, each with its own strengths and weaknesses. The idea of optimality implies that, given a particular context or set of constraints, there is a best-suited strategy that maximizes the outcome while minimizing costs, such as cognitive effort, time, or resources.

To continue the previous decision-making story of Mona, she decided to major in biology at the beginning of her freshman year. She is confronted with the challenge of deciding whether to continue her major in biology or switch to a different field, such as psychology. She was passionate about the subject and initially believed that pursuing a career in biological research was the perfect path for her. Over time, however, she found that her enthusiasm for the subject waned, and she started to develop a keen interest in psychology instead.

Despite her growing disinterest in biology, Mona feels compelled to continue her major, as she has already invested a significant amount of time, effort, and money into her biology coursework. In addition, she worries that if she switched majors, she would have wasted all her previous investments in her biology classes. This example demonstrates the sunk cost fallacy at play in Mona's decision-making. Rather than objectively evaluating her situation and making a decision based on her current interests and potential career paths, Mona allows her past investments in her biology major to unduly influence her decision to stick with the program.

To overcome the sunk cost fallacy and make a more optimal decision, Mona would need to engage in cognitive control and exert cognitive effort to evaluate her situation objectively. This may involve weighing the benefits of switching to a psychology major against the potential costs of doing so, such as taking longer to graduate or needing to catch up on coursework in her new field. By considering these factors and separating her past investments from her current interests and goals, Mona could make a more informed and rational decision about her academic path.

Examining major life decisions, such as selecting a college major, can reveal how people navigate the myriad influencing factors, and it is essential to also investigate the relationship between cognitive effort and optimal decision-making in simpler, everyday scenarios. By narrowing our focus to more straightforward decision-making situations, we can closely examine how cognitive processes like model-free (MF) and model-based (MB) learning interact and influence our choices.

Imagine Mona has a busy schedule balancing her academic and social life. On a typical day, she needs to make decisions about how to commute to school and manage her time effectively. Drawing from her past experiences, she knows a routine route to her class that takes 15 minutes by car. Mona typically relies on her model-free (MF) decision-making, where she uses previously learned values to guide her choices.

However, today is different. Mona’s MB decision-making comes into play as she considers the possibility of a football game event affecting traffic. She knows that taking a local route may add an extra 5 minutes to her commute, but it could save time by avoiding traffic on the highway. The MB approach allows her to simulate different scenarios and strategically plan her actions based on her internal model of the environment.

Weighing the pros and cons, Mona decides to take the local route, hoping that the time saved will outweigh the additional time it potentially adds. In this scenario, Mona's decision-making involves both the MF and MB approaches, allowing her to adapt to changes in her environment and make more informed choices in her daily life.

1.7.1 *Theories of Decision-Making*

The normative theory of decision making is the Expected Utility Theory, which assumes rationality in the decision-maker. According to this theory, individuals select the option with the highest expected utility, which is calculated by multiplying the probability of an outcome by the utility of that outcome. This model has been widely used in economics and decision-making research, but it has also been criticized for its assumption of perfect rationality, which is not always realistic.

Prospect Theory, proposed by Daniel Kahneman and Amos Tversky, provides a more accurate description of human decision-making by accounting for cognitive biases and deviations from perfect rationality. This theory posits that people evaluate potential gains and losses relative to a reference point, and they are more sensitive to losses than gains. Prospect Theory also incorporates the concept of probability weighting, where people tend to overestimate the likelihood of rare events and underestimate the probability of common events.

Reinforcement learning is a computational approach to understanding decision-making based on the idea that individuals learn from the outcomes of their actions. Model-free (MF) learning is based on trial-and-error, where decision-makers update their value estimates based on the rewards or punishments they receive. MB learning, on the other hand, involves building a mental model of the environment and planning actions based on that model. Both MF and MB learning processes contribute to decision-making, and their relative influence depends on factors such as task complexity and individual preferences.

While both MB and MF approaches contribute to decision-making, the precise balance between these systems and their cognitive mechanisms in shaping distinct decision-making strategies remains a subject of ongoing debate. In the realm of reinforcement learning (RL), MB and MF approaches exhibit distinct characteristics: MB-RL focuses on creating an environmental model, allowing the agent to plan actions by simulating potential outcomes, whereas MF-RL eschews environmental modeling in favor of trial-and-error learning based on received rewards. The choice between MB and MF-RL hinges on the specific task and available information, with each approach offering unique advantages and limitations.

Cognitive psychology literature often discusses MB and MF in terms of their underlying cognitive processes and implications for decision-making. MB is believed to rely on cognitive processes such as working memory and planning, while MF depends more on simple associations between actions and rewards. Studies have demonstrated that MB is more effective in stable and predictable environments, while MF is better suited to changing or uncertain conditions. Moreover, research has revealed that the balance between MB and MF shifts throughout development and aging, and individuals with certain neurological disorders may exhibit a preference for one learning style over the other.

Doll et al. (2015) provide critical insights into the interplay between MB-RL and MF-RL approaches in decision-making strategies. They pointed out that the brain's multiple memory systems, specifically the declarative and procedural memory systems, serve as substrates for the distinct decision systems. Declarative memory, which involves conscious recollection of facts and events, is associated with MB-RL, as it enables the construction of internal models of the environment and the planning of actions based on those models. In contrast, procedural memory, which entails learning habits and skills, is linked to model-free reinforcement learning, where decisions are guided by learned associations between actions and outcomes. Doll et al.'s findings further emphasize the interdependence of these two memory systems and decision-making strategies, highlighting the complex and adaptive nature of human cognition. Their work underscores the necessity of considering the contributions of multiple memory systems when investigating the balance between model-based and model-free approaches in decision-making.

Understanding strategy selection is essential for improving decision-making processes, as it allows individuals to adapt their approach according to the specific demands of a situation. By identifying the optimal strategy, individuals can enhance the quality of their decisions, ultimately leading to better outcomes in various domains, such as personal, professional, or educational settings. Moreover, understanding the factors that contribute to optimal strategy selection can provide insights into the cognitive mechanisms underlying decision-making, allowing researchers to develop better models and theories of human cognition.

Instance-Based Learning (IBL) is a computational framework that emphasizes the role of memory in decision-making. In IBL, decision-makers encode experiences as instances in their memory and retrieve these instances to inform future choices. This approach captures the dynamic nature of decision-making and offers a mechanistic understanding of how individuals learn from their experiences and optimize their actions based on rewards and punishments.

In addition to its theoretical importance, understanding strategy selection has practical implications for designing interventions and training programs aimed at improving decision-making. By identifying the factors that facilitate or hinder the selection of optimal strategies, interventions can be tailored to target specific weaknesses or promote the use of more effective strategies. For example, decision-making training could focus on helping individuals recognize the contexts in which certain strategies are most effective, or on providing tools for overcoming cognitive biases that impede optimal strategy selection. Ultimately, a deeper understanding of strategy selection in decision-making has the potential to significantly enhance the quality of decisions made in various aspects of human life.

1.7.2 *Factors that influence optimal decision-making*

Optimal strategy selection in decision-making is influenced by various interrelated factors. One important aspect is cognitive factors, which include cognitive abilities, working memory capacity, and cognitive biases and heuristics. These elements can significantly shape an individual's ability to process information, think critically, and employ effective decision-making strategies.

Task characteristics also play a role in optimal strategy selection. Factors like task complexity, time constraints, and uncertainty can affect the choice of strategies based on the specific demands of a given task. Depending on the task's requirements, individuals may need to adapt their decision-making strategies to achieve the best results.

Additional components involve emotional factors, such as emotional regulation and affective forecasting. These factors impact strategy selection by determining how well individuals manage their emotions and predict their emotional responses to potential outcomes. This, in turn, can influence their decision-making processes and the strategies they choose to employ.

Lastly, contextual factors, encompassing environmental conditions and social and cultural influences, can shape decision-making processes and the selection of optimal strategies. These external factors can have a considerable impact on individuals' decision-making abilities and the strategies they ultimately choose to use.

In summary, optimal strategy selection in decision-making is influenced by a variety of factors. Cognitive factors, such as cognitive abilities, working memory capacity, and cognitive biases, play a crucial role in determining the most effective decision-making strategies. Emotional factors, including emotional regulation, also impact strategy selection by shaping how individuals process and respond to their emotions during decision-making. Task characteristics, such as complexity, time constraints, and uncertainty, influence the choice of strategies based on the specific demands of a task. Finally, contextual factors, including environmental conditions and social and cultural influences, can affect decision-making processes and the selection of optimal strategies. By considering these factors, individuals can better understand and adapt their decision-making strategies to achieve more optimal outcomes.

Chapter 2. COMPUTATIONAL FRAMEWORKS

In this chapter, I will discuss key computational frameworks that researchers have used to explain human decision making. Although many such frameworks exist, I will focus here on explanatory theories that can be naturally interpreted in terms of cognitive process – that is, theories that exist at Marr's "algorithmic" level (Marr, 1982) and clearly articulate the computations associated with making a decision. Specifically, I will focus on Reinforcement Learning (RL), which is currently the dominant theory to explain experiential decision-making in humans and in animal models. In addition, I will examine the principles and applications of a related framework, Expected Value of Control (EVC) theory. Furthermore, I will present the ACT-R cognitive architecture, a framework for simulating human cognitive processes and, in particular, the mechanisms of memory encoding and retrieval. Finally, I discuss the relationship between RL, EVC theory, and ACT-R, focusing on their potential integration and the broader implications of combining these potent computational approaches for modeling human cognition.

One key advantage of employing computational modeling in cognitive psychology is their ability to capture individual differences in learning and decision-making strategies. By fitting precise algorithmic models to empirical data, researchers can measure the values of specific parameters that affect a person's ability to carry out the mental steps associated with a decision – for example, the probability of correctly remembering the outcome of a previous choice. This insight can help reveal the neurobiological substrates associated with different decision-making strategies, contributing to a better understanding of the neural basis of cognition.

2.1 Reinforcement Learning

Reinforcement Learning (RL) in computational modeling shares some similarities with the Reinforcement Theory of motivation in psychology, but they are not entirely identical. While taking inspiration from early versions of Reinforcement Theory (such as the Rescorla-Wagner rule: Rescorla & Wagner, 1975) the field of RL has mostly developed within the artificial intelligence community to solve practical problems by learning optimal actions from environmental feedback. As it turns out, this approach has led to important insights into the behavioral and neural mechanisms that underpin learning and decision-making (e.g., Schultz, Dayan, and Montague, 1997).

Under the RL paradigm, an agent explores the environment at time (*t)*, evaluates the current state (*s*), selects an action (*a*) from multiple possible actions, and receives feedback, in the form of a scalar reward signal (*r)*, from the external world. As time progresses, the influence of reward on current action decisions diminishes. The goal of the agent is to maximize the sum of the foreseeable future reward. Figure 2.1 illustrates how the RL agent interacts with the environment by receiving rewards and initiating actions.

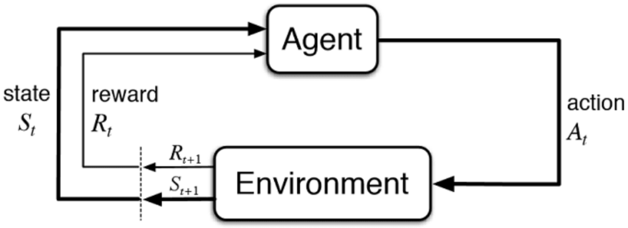

Fig 2.1. The action-reward feedback loop of a generic Reinforcement Learning model.

Interestingly, the concept of RL, rooted in trial-and-error learning, closely aligns with the adaptive learning process in humans and other animals. This process involves using prior knowledge to evaluate current states, decide on the optimal sequence of actions, and improve the outcomes of choices. Two prominent approaches within RL are the model-free (MF) and model-based (MB) methods. MF-RL relies on the agent learning from direct experience, using techniques such as temporal difference learning (Sutton, 1988) to estimate state-action values (technically known as Q-values) without explicitly modeling the environment's dynamics.Thus, in this family of RL models, agents simply learn the reward values associated with each action. For example, in the RL algorithm known as SARSA, an agent updates the value of a state-action pair, Q(s,a) according to the following rule:

$$Q_{new}(s', a') = Q_{old}(s, a) + \alpha\delta_{i,t} \text{ where } \delta_{i,t} = r + \gamma Q(s'_t, a'_t) - Q(s_t, a_t)] \quad (2.1)$$

α is the learning rate, *r* is the reward, and Q(s’, a’) is the value of the next action that the agent is about to take. Because it combines information from two consecutive states (s and s’) and actions (a and a’), SARSA can learn complex sequences of behavior. However, it does so in a “habitual” way, mindlessly following the series of actions that have, in the past, yielded the most rewards.

In contrast, MB agents build an internal model of the environment, allowing the agent to plan and reason about future outcomes and rewards. While MF approaches tend to look extremely similar to one another, MB approaches can be very different from each other, depending on the way they learn from the environment and plan their moves. For example, the

MB model that will be discussed in Chapter 5 explicitly stores the transition probability P(s,a,s’) between two consecutive states, and uses this information to calculate the best action:

$$Q(s,\ a)\ =\ r\ +\gamma \sum^{s'} P(s' \mid s,\ a) \cdot max[Q(s',\ a)] \tag{2.2}$$

While MF methods are computationally less demanding and more reactive, MB methods enable better long-term planning and can adapt more rapidly to environmental changes. Both MF and MB approaches have been used to explain various aspects of human decision-making (Daw et al., 2011; Otto et al., 2013; Doll et al., 2015), and their integration in hybrid models can provide a more comprehensive understanding of cognitive processes.

2.2 THE EXPECTED VALUE OF CONTROL (EVC) THEORY

RL models do not consider either the costs associated with each action or the motivation an agent might have to achieve a given goal — in fact, RL models have no goal whatsoever, if not to maximize their future rewards. In contrast, biological agents have explicit goals and objectives that guide their behavior, which can be influenced by both intrinsic motivations and external factors. Importantly, biological agents also have to consider the costs associated with each action, which can involve cognitive and physical effort, as well as potential risks or negative consequences.

This is where the Expected Value of Control (EVC) theory comes into play. Proposed by Amitai Shenhav, Matthew Botvinick, and others, EVC theory provides a framework to understand how biological agents decide which cognitive control to exert in response to their environment. It argues that the brain selects actions by calculating their expected value, which balances the potential rewards against the estimated costs of control, such as cognitive effort (Shenhav 2013; Shenhave 2017; Gergelyfi 2015, Kool et al., 2018). The Expected value of Control (EVC) model, not only to account for which actions an individual may choose among many alternatives but also describe how much cognitive effort one would allocate to the chosen action, as shown in Figure 2.2.

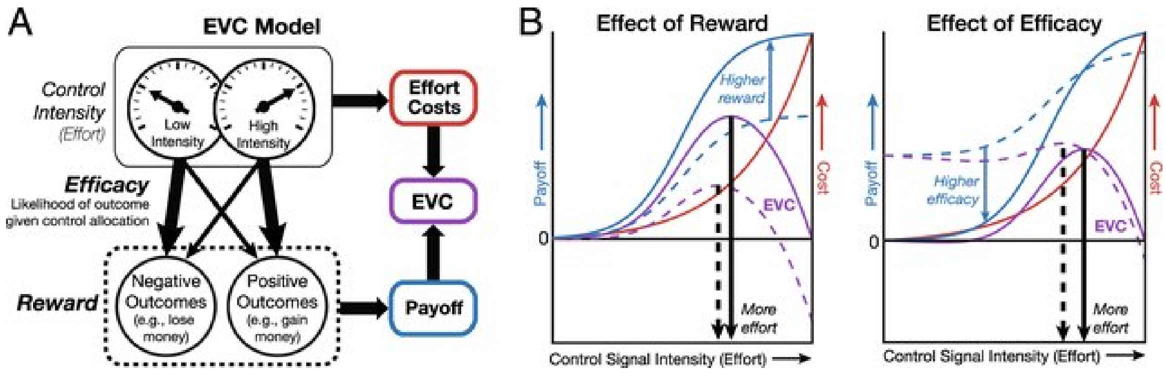

Fig 2.2. The Expected Value of Control Model adapted from Shehav et al. 2017. *(A) The relationship between costs, payoff, and expected value of control (EVC). (B) The expected value of control (EVC) model predicts that control should increase with expected reward and efficacy.*

According to the EVC model, the expected value of control, EVC, is determined by the expected reward and efficacy of the task. Expected reward indicates the expected outcomes of achieving the goal (i.e. monetary incentives), and efficacy refers to how likely the goal will be achieved by allocating a certain amount of control. Take real-life bicycle racing as a comparable example of physical effort. To ride faster, the rider has to exert more force on pedaling (physical control). The harder and faster he/she pushes on the pedal, the more likely they are to win the race, getting positive outcomes (better outcome). However, such effort is associated with intrinsic costs and fatigue. If the rider has high efficacy of the positive outcome, and the possibility of winning the race is high, he/she may exert even more effort to overcome the effort and to achieve the goal. On the contrary, if the efficacy is low, the rider might choose to preserve the resources by not investing much effort in the task.

Computationally, the EVC model specifies that cognitive effort is allocated based on two dimensions: 1) identify what signal to attend; and 2) determine the intensity of control(how much effort to allocate on, compared to default level). The key assumption of this model is that an intrinsic cost would be associated with higher control intensity. The expected value of control (EVC) given an outcome and state is calculated, state and is calculated as below (Eq 2.3), where *signal* indicates the control signal, represented by which signal the agent focuses on, and the intensity of the task; *state* indicates the current situation, represented by task difficulty, motivational state etc; *outcome* is represented by the reward received by executing the action, such as correct/incorrect feedback.

$$EVC(signal,\ state) = \sum_{i} Pr(o|signal,\ state) * Value(o_i) - Cost(signal) \qquad (2.3)$$

Inherent from the RL algorithm, an agent should maximize the rate of reward which is determined by the two dimensions discussed above. Taking the Stroop task as an example (shown in Figure 2.2), which involves incongruent stimuli, typically color words printed in different colors, the agent needs to pay attention to two types of signals: (1) word meaning, and (2) word color, making it a classic cognitive psychology experiment that measures cognitive control and interference in response to conflicting information. Depending on the congruency of the trial (congruent condition: “**BLUE**”; incongruent condition: “**BLUE**”), the intensity of the signal varies. For instance, when seeing a stimulus “**BLUE**”, lower intensity of control (also called “effort”) is required on naming color, while when seeing a stimulus “**BLUE**”, higher intensity of control is required for naming color. Every trial will lead to either correct or incorrect feedback, which determines the final payoff an agent would receive. By adjusting different amounts of control, an agent calculates the corresponding EVC value using Equation 1.

To answer the question of what motivation is, the EVC model describes motivation as a cost-benefit decision-making analysis. To answer the question of how motivation affects behavior, the EVC model argues that after evaluating cognitive effort costs, potential payoffs, and efficacy, the optimal cognitive effort allocated to initiate a behavior is calculated based on the EVC value. Higher payoffs will increase the reward curve (blue curve) in Figure 2.2, which makes the EVC curve shift to the right, and the optimal cognitive effort increases. Higher task difficulty suggests lower efficacy, because the probability of reaching goals becomes smaller. Thus, the payoff curve tends to be flat when efficacy is low, and even less effort will be invested. This explains why the task is too hard, people may feel frustrated and quit the task.

This EVC model integrates goal achievement, decision-making, and effort evaluation into motivation, attempting to account for how each choice is being made depending on expectancy and value. This model also accounts for the relationship between effort and motivation, which is overlooked in other major theories of motivation. It not only provides a qualitative definition of motivation but also quantifies motivation in terms of two variables: expectancy and value. Admittedly, the shape of the cost-effort relationship requires further examination. Kool and Botvinick (2018) proposed that the cost-effort function may be in line with the concave shape of the utility function from the economic labor supply theory.

The EVC theory and RL can be linked in the context of decision-making processes. For instance, EVC theory can be viewed as a higher-level cognitive framework that guides the allocation of cognitive resources during decision-making. RL, in turn, can be seen as a mechanism through which these resource allocation decisions are implemented and refined based on feedback from the environment. Thus, while EVC theory provides a normative framework for understanding how cognitive control should be allocated optimally, RL provides a mechanism for how these optimal control policies might be learned and updated over time based on experience. This combination offers a comprehensive view of decision-making processes, encompassing both the strategic allocation of cognitive resources and the learning processes that allow these strategies to be refined and optimized.

2.3 ACT-R ARCHITECTURE

Adaptive Control of Thought-Rational, or ACT-R for short, is the most well-known and rapidly growing cognitive architecture. It was developed by John Anderson and his colleagues to simulate and explain a wide variety of cognitive processes in neuroscience and psychology, such as perception, memory, language, learning, and decision-making. It was inspired by the concept that human cognition can be represented as a collection of symbolic productions, dictionary-like memory, and functioning on a set of symbolic representations. ACT-R assumes that human thinking is a rational system that makes decisions by selecting the action that maximizes expected utility. Figure 2.3 demonstrates the basic components and organization of ACT-R architecture.

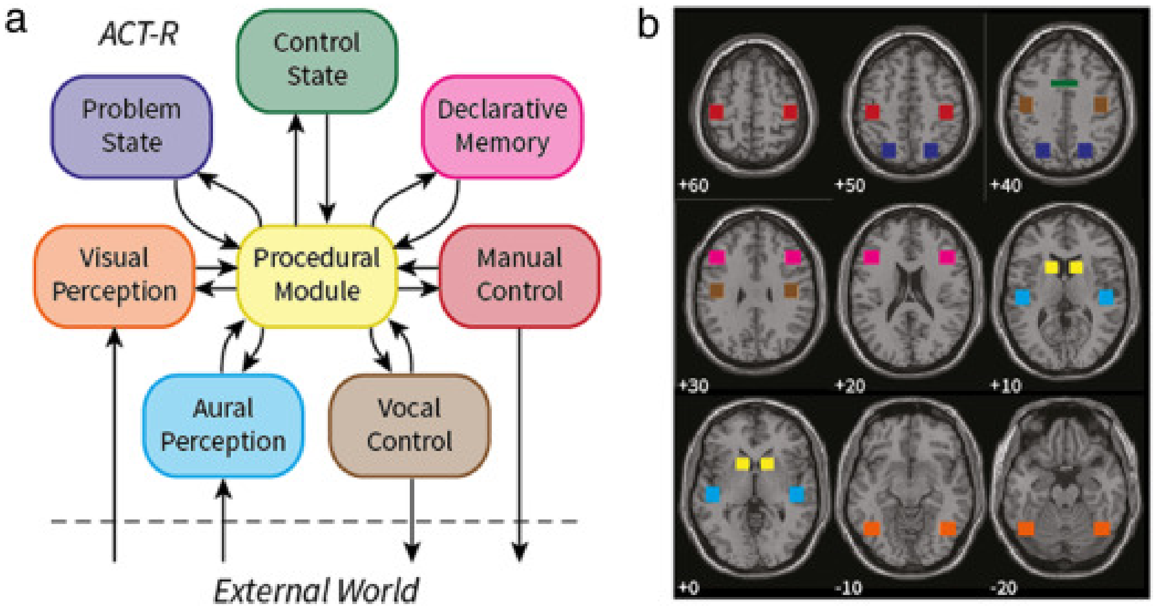

Fig 2.3. The illustration of ACT-R architecture. *(A) The abstract representation of ACT-R systems with 8 major components. (B) The neural mappings of eight modules in the human brain.*

In ACT-R, the concept of knowledge is broken down into its two primary components: chunks and production rules, as shown in Figure 2.4. Memory, either semantic or episodic, may be stored in a form similar to a vector called a chunk. A production rule, sometimes known simply as production, is a fundamental action unit that expresses procedural knowledge in the form of an “IF-THEN” conditional statement. The interaction between productions and chunks is mediated by a collection of modules that stand in for a variety of cognitive processes. The ability to replicate a variety of cognitive tasks is made possible by ACT-R’s of both a declarative memory system and a production rule system. This involves the capacity to represent and manage ambiguous information, the ability to learn from previous experience, and the ability to strike a balance between exploring new choices and using the knowledge that has been learned.

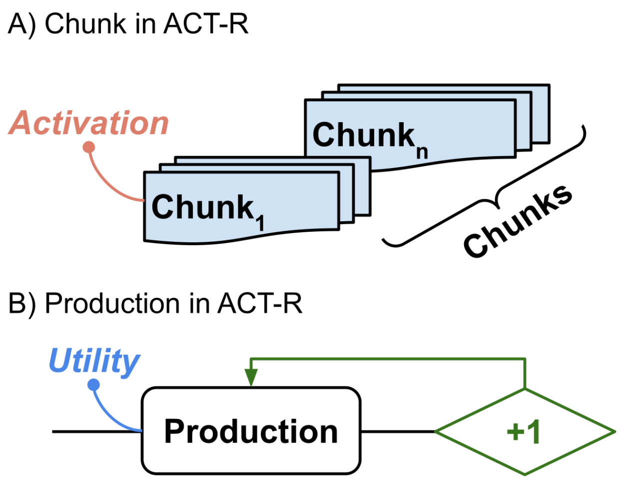

Fig 2.4. The illustration of the basic construction element in ACT-R. *A) Chunk. B) Production*

Activation. Each chunk is associated with a scalar value, called activation, which represents the log odds of a chunk being needed in the future (Anderson, 1998). The activation of a chunk in declarative memory is calculated using the following equation (Eq 2.4):

$$A_i = B_i + \varepsilon \qquad (2.4)$$

- where A_i is the activation of chunk i;
- B_i is the base-level activation, which reflects the recency and frequency of practice of the chunk i;
- ε is the noise value, which usually captures the inherent uncertainty and variability in human memory retrieval.

The equation describing learning of base-level activation for a chunk i is shown in Eq 2.5:

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \qquad (2.5)$$

- where n is the number of presentations for chunk i;
- t is the time since the jth presentation;
- d is the memory decay parameter.

The activation of a chunk consists of two major components, base-level activation, and spreading activation, reflecting both long-term usefulness and short-term relevance. This allows the cognitive system to adaptively prioritize memory retrieval. The decay parameter (d) captures the forgetting process, as older memories have lower activation levels due to the passage of time. The noise term (ε) accounts for the variability in human memory performance, ensuring that the model can generate realistic predictions of human behavior. The activation equation provides a quantitative and testable framework for understanding human memory retrieval, allowing researchers to study the factors that influence memory activation and the strategies employed by individuals during decision-making.

In ACT-R, the probability of retrieving a chunk is determined by activation value, noise, and threshold, as shown in Eq 3. As long as the activation exceeds the threshold, higher activation usually suggests that the memory is more likely to be retrieved, $Probability$ is closer to 1. τ and s are two parameters scale probability in different ways. As τ becomes greater, the probability decreases. As s increases, the recall sensitivity becomes a slow sigmoid curve.

$$Probability = \frac{1}{1 + e^{\frac{\tau - A_i}{s}}} \qquad (2.6)$$

- where A_i is the activation of chunk i;

- τ is the retrieval threshold
- s is the parameter that controls the sensitivity of recall to changes in activation

One of the important and prominent features of ACT-R is its ability to estimate the retrieval latency based on the activation of memory. Eq 5 below shows how ACT-R calculates the response time required to retrieve a memory. Usually, the higher the activation level of a chunk (A_i), the shorter its retrieval latency. This relationship is governed by an exponential function, which implies that chunks with higher activation levels are retrieved more quickly than those with lower activation levels.

$$Time = F \cdot e^{-A_i} \qquad (2.7)$$

- where A_i is the activation of chunk i;
- F is the latency factor parameter

The retrieval latency equation demonstrates that memory retrieval is a dynamic and time-dependent process. It recognizes the fact that human cognitive processes have time constraints, making the model more cognitively plausible and realistic in accounting for the human brain. The relationship between activation and retrieval latency allows the cognitive system to prioritize the retrieval of more relevant and useful information, as chunks with higher activation levels are retrieved faster. The latency factor (F) allows researchers to control the speed of retrieval and to account for individual differences in retrieval speed or cognitive processing abilities. The retrieval latency equation provides a quantitative approach to understanding the dynamics of memory retrieval and the time it takes to process information during decision-making. This facilitates the study of time-sensitive cognitive tasks and the development of more accurate models of human behavior.

Utility. Each production is associated with a utility value that represents the expected future rewards associated with the execution of that production (and is analogous to Q-values in Reinforcement Learning: Niv, 2009). In ACT-R, only one chunk can be retrieved and only one production can be fired at a time. Competing chunks are selected based on their activations. Similarly, several candidate production rules are competing with each other, only one of which is selected based on its utility. Utilities are learned through experience. At any time point t, the utility U of production p is calculated based on Temporal Difference Reinforcement Learning using Eq 6, where α denotes the learning rate, R_t denotes the reward the production received at time t; s denotes the noise parameter.

$$U_t = U_{t-1} + \alpha(R_t - U_{t-1}) + s \qquad (2.8)$$

The utility equation captures how valuable a production rule is in the context of a specific task or decision-making process, quantified using utility value U at a certain time t. The utility is a combination of the expected probability that a production rule will lead to the expected outcome and the expected reward that the environment delivered when applied successfully. Like most RL Q-Learning, the parameter α controls how fast the system can learn the reward outcome association, thereby influencing the selection of production rules during decision-making.

The utility equation provides a mechanism for selecting the most appropriate production rule based on the current context and the expected outcomes, enabling adaptive decision-making in the cognitive system. The utility equation acknowledges the role of rewards and gains in shaping human decision-making, which is consistent with the principles of reinforcement learning and motivation. With the help of these parameters, researchers can model individual differences in the weighting of gains and probabilities, providing a more flexible and quantitative approach to understanding the role of procedural memory in human decision-making strategies.

2.4 SUMMARY

Above, I have explored three influential frameworks in modeling decision-making: Reinforcement Learning (RL), Expected Value of Control (EVC) theory, and the ACT-R cognitive architecture. Each of these frameworks offers unique perspectives and contributions to our understanding of the complex processes underlying human decision-making. In this summary, I will synthesize these frameworks and highlight their interrelationships in the context of modeling decision-making.

Reinforcement Learning (RL) is a computational framework that focuses on learning how to act optimally by receiving feedback from the environment. At its core, RL is centered around estimating action values (Q-values) and using them to determine the best sequence of actions that maximize cumulative rewards. RL has been extensively studied in understanding the neural basis of decision-making and has been widely applied in various domains, including psychology, neuroscience, and artificial intelligence. The Expected Value of Control (EVC) theory extends the RL framework by incorporating the trade-offs between expected rewards and costs associated with cognitive control. EVC theory provides a high-level understanding of the neural mechanisms underpinning motivation, particularly in the context of the dorsal ACC. While EVC theory offers valuable insights into the neural basis of motivation, it remains abstract and does not provide specific mechanisms for computing costs and rewards or predicting motivational shifts in task performance.

The ACT-R cognitive architecture bridges this gap by offering a fine-grained representation of human cognition and incorporating goal-related constructs. Specifically, the relationship between the Expected Value of Control (EVC) theory and ACT-R can be elucidated by examining their respective roles in modeling human cognition and motivation. EVC theory, with its focus on the trade-offs between expected rewards and costs, provides a high-level framework for understanding the neural mechanisms underpinning motivation, particularly in the context of the dorsal ACC. However, EVC theory remains abstract and does not offer specific mechanisms for computing costs and rewards or predicting motivational shifts in task performance. This is where the ACT-R cognitive architecture, with its fine-grained representation of human cognition, comes into play.

ACT-R allows for the incorporation of motivation-related constructs such as intrinsic motivation, goal competition, emotion, mental fatigue, and depression. The activation of chunks and the utility of production rules in ACT-R can be linked to EVC theory, enabling a more detailed understanding of how motivation influences cognitive processes and control. While EVC theory provides a foundation for understanding the neural basis of motivation and its role in decision-making, the integration of ACT-R allows for the exploration of the dynamic interplay between motivation and cognitive control. By connecting the Goal module in ACT-R to the dorsal ACC and rewards, researchers can develop a more comprehensive model that aligns with established neuroscience findings and is compatible with the EVC model.

Furthermore, by revisiting earlier versions of the ACT-R architecture that featured explicit goal values and cost-benefit analyses, researchers can develop a more unified framework that incorporates the reinforcement learning-based utility mechanisms of the current ACT-R model. This updated framework would bridge the gap between the high-level abstractions of EVC theory and the detailed mechanisms of ACT-R, ultimately providing a more complete understanding of the relationship between motivation and cognitive control.

In conclusion, the relationship between EVC theory and ACT-R lies in their complementary roles in modeling human cognition and motivation. EVC theory offers a high-level perspective on the neural mechanisms underlying motivation, while ACT-R provides a detailed account of cognitive processes and their interactions. By integrating these two frameworks, researchers can gain a more comprehensive understanding of the complex interplay between motivation and cognitive control, leading to novel insights and applications in psychology and neuroscience.

Chapter 3. STUDY 1: A MODEL OF OPTIMAL EFFORT ALLOCATION IN A COGNITIVE TASK [1]

3.1 ABSTRACT

This study outlined a general framework of goal-oriented decision-making strategy in ACT-R that is consistent with EVC theory (Shenhav et al., 2017), where the optimal strategy an individual adopts to allocate cognitive effort is represented as the process of multiple goals competing with one another. Most importantly, in this study, we proposed a plausible cognitive framework demonstrating that this process is determined by a mechanism similar to memory retrieval in ACT-R. This innovative paradigm enables us to first consider cognitive allocation decision-making techniques in terms of memory. To demonstrate this, we developed two models inside the ACT-R framework, demonstrating that ACT-R is capable of picking the best approach by considering costs and benefits in accordance with EVC Theory.

3.2 BACKGROUND

Observable behavior in cognitive tasks is affected by the degree to which a participant puts effort into the task. The driving force behind this effort allocation is usually called motivation and represents a significant obstacle in properly inferring individual characteristics from observations. For example, a participant performing poorly in an N-back task might be poorly motivated to perform the task, rather than having limited working memory capacity. Despite its importance, motivation is rarely modeled or accounted for in cognitive models. In this

---

[1] Parts of this chapter have been presented at the 2021 International Conference on Cognitive Modeling, and are currently under review for publication at *Topics in Cognitive Science*.

paper, we outline a theory of motivation implemented in the ACT-R cognitive architecture and demonstrate its application.

To understand motivation from a cognitive modeling perspective, it is necessary to clarify the definition and relationships between several important constructs. *Motivation* is not directly observable. It is usually described as a driving force or invigorating impact on behavior or cognition that initiates a goal-oriented behavior. That is to say, we can only infer one's motivation from his behavior and cognition. *Effort* refers to how many cognitive resources one would allocate to a particular activity in order to achieve the goal. According to Inzlicht, Shenhav, and Olivola (2018), Motivation specifies both direction and intensity of goal-oriented behavior, while effort only indicates the intensity of any possible action, without reference to any goal. Finally, *Demand* differs from *Effort* in that it is the descriptive property of the task or environment, while *Effort* indicates the subjective magnitude of the force that an individual might apply toward the environment; in other words, demand is a function of the tasks while effort is a function of the individual. Other cognitive states such as mental fatigue, curiosity, and high arousal may interact with motivation in certain ways to have crucial impacts on learning, memory, and other cognitive control functions therefore, cognitive modeling gives us a unique opportunity to parse apart the specific effect of motivation alone.

The Expected Value of Control Theory. Although several attempts have been made to capture motivation within a computational framework (e.g., Niv, 2007), the present dominant theory is the Expectancy Value Theory. It was first informally proposed by Voom in the 1960s and recently formalized as the Expected Value of Control (EVC) model by Shenhav et al. (2017). The EVC model assumes that individuals would evaluate cost-benefit tradeoffs in order to maximize gains and minimize costs in deciding how much cognitive effort one would allocate to the chosen action, as shown in Figure 3.3 (A, B).

According to the EVC model, the expected value of control is determined by the expected reward and efficacy of the task. The expected reward indicates the expected outcomes of achieving the goal (e.g., monetary incentives) and efficacy refers to how likely the goal will be achieved by allocating a certain amount of control and expending a certain amount of effort. For example, a student could put more effort (i.e., allocating more time) into studying for a test to increase the probability (efficacy) of getting a good grade (reward). Computationally, the EVC model specifies that cognitive effort is allocated based on two dimensions: 1) identify what signal to attend ; and 2) determine the intensity of control(how much effort to allocate on, compared to default level). A key assumption of this model is that intrinsic costs would be associated with higher control intensity. At first glance, the definition of intrinsic (cognitive) cost and control intensity may appear similar, yet there is a fundamental difference between “control intensity” and “intrinsic cost”. The terms “cognitive control”, “cognitive effort”, and “effort allocation” refer to the “decision” of how many cognitive resources to put into a task, after taking reward, cost, and many other factors into account. In contrast, intrinsic (cognitive) cost refers to the base level of cognitive resources needed in order to complete a task or achieve a goal, which is closer to a trait, characteristic, or capability of an individual and, more importantly, is supposed to be independent of the environment and task. For example, an individual with lower working memory capacity would experience higher intrinsic costs in the N-back task than an individual with higher capacity; however, the former might allocate more cognitive control, and thus put more cognitive effort into the task and outperform the latter.

At the neural level, the translation between the expected value (i.e., the difference between expected rewards and costs) and corresponding effort allocation is mediated by the dorsal ACC, a region that is known to play a critical role in linking adjustments in performance (Botvinik et al., 1999) with task feedback (Holroyd et al., 2004), error learning (Yeung et al., 2004) and with expected rewards (Adcock et al., 2006).

Thus defined, EVC theory is an elegant, comprehensive, but highly abstract framework: it does not provide a direct mechanism by which costs and rewards are computed and associated to specific cognitive steps, nor it does make specific predictions about how motivation would precisely shift how an individual performs a task. To do so, we need a more fine-grained and detailed theory of human cognition. One such prominent theory is the ACT-R cognitive architecture (Anderson, 2007).

ACT-R is the most prominent and successful cognitive architecture in psychology and neuroscience (Kotseruba and Tsostsos, 2020). Surprisingly, despite the high relevance of motivation to other cognitive functions and the apparent potential and an ACT-R model of motivation, the interaction between motivation and cognitive control has been largely overlooked in ACT-R literature. Several modeling attempts have been made in order to incorporate motivation-related constructs into ACT-R, such as intrinsic motivation (Nagashima et al., 2020), goal competition (Herlambang, Taatgen & Cnossen, 2020), emotion (Smith et al., 2021), mental fatigue (Herlambang et al., 2021; Halverson et al., 2021), and depression (van Vugt and van der Velde, 2018).

As noted in Chapter 2, In ACT-R, knowledge is represented in two fundamental formats: chunks and production rules. A Chunk is a vector-like structure that stores semantic or episodic memories. A Production rule (or simply production) is a basic action unit that represents procedural knowledge as an “IF-THEN” conditional statement. Productions and chunks interact through a set of modules that represent different cognitive processes. For example, a Visual module encodes visual information as chunks, and a Motor module transforms chunks into motor outputs. Most critical to this paper are the Goal module (holding current goal information), the Declarative module (storing all declarative memories and managing their availability for retrieval), and the Procedural module (maintaining, updating, and selecting productions).

Each chunk is associated with a scalar value, called activation (see Chapter 2.3, Eq 2.4, Eq. 2.5), which represents the log odds of a chunk being needed in the future (Anderson, 1998). Similarly, each production is associated with a utility value that represents the expected future rewards associated with the execution of that production (and is analogous to Q-values in Reinforcement Learning: Niv, 2009). Utilities are learned through experience. At any time point t, the utility U of production p is calculated based on Reinforcement Learning using Eq 3.1, where α denotes the learning rate, R_t denotes the reward the production received at time t; s denotes the noise parameter (See Chapter 2.3, Eq 2.8).

$$U_t = U_{t-1} + \alpha(R_t - U_{t-1}) + s \qquad (3.1)$$

In ACT-R, both rewards and costs are represented in time units. For instance, if the model fires a production P_1 at t_1 and it receives a reward R at t_2. The utility learning mechanism discounts the reward by the amount of time that has passed since the production has fired. Specifically, the received amount of reward is:

$$R_{received} = R_{delivered} - (t_2 - t_1). \qquad (3.2)$$

It should be noted that, in ACT-R, the above-mentioned Goal module is putatively associated with the dorsal ACC (Anderson, 2007), but has no relationship to rewards and is, in fact, used only as a way to add additional information to select between competing productions. This violates established findings in neuroscience and is incompatible with the EVC model. It is also a major departure from early versions of the ACT-R architecture (e.g., Anderson & Lebiere, 1998), in which goals were associated with specific values, and values were explicitly used to rank productions on the basis of a cost-benefit analysis. This older framework was, in principle, much more compatible with EVC theory, as it explicitly selected strategies based on a cost-benefit analysis of goal values and the time needed to achieve them. In this older framework, productions were selected not on the bases of utility but on the bases of an estimated quantity expressed as $pG - C$, where p is the probability of achieving that goal, G is the goal's value, and C is the time cost needed to achieve it (Anderson & Lebiere, 1998).

One of our objectives is to propose a framework that conserves the current RL-based utility mechanisms but connects it with explicit goal values, re-introducing some of the most desirable features of the previous implementations.

The goal of this study is to outline a general framework of goal-oriented motivation in ACT-R that is consistent with EVC theory and can be implemented and deployed in any ACT-R model. This is implemented by assigning to the current Goal chunk a special motivation slot that contains a numeric value M. Once the goal is achieved, M is interpreted as the amount of reward delivered in the end. At that moment, M is automatically translated into the R_t value that is used in Eq 1 and propagates back to previous productions. Because in ACT-R, as noted, rewards are represented in time units, the value M can be interpreted in two ways: as the subjective value associated with reaching a goal, as the maximum amount of time the model is willing to spend on a particular goal. The first interpretation is consistent with the current interpretation of the reward value R_t, while the second is consistent with the original interpretation of the goal value G in previous versions of ACT-R. By incorporating this mechanism, the Goal buffer is not only a passive recorder of task status, but an active power behind adaptive behaviors, and as the interface between the goal-directed behavior of the architecture, which relies on explicit chunk representations, and the reinforcement learning mechanisms that affect the selection of productions. Because rewards are already affected by the passing of time, our model does not require the time scaling mechanism of Harlembang et al. Crucially, our model also attempts to account for where the intrinsic reward R_t comes from, and how motivation value M alters one's behavior by affecting the neurocomputation of expected reward and effort.

Our framework takes a different approach than that taken by Herlambang, Taatgen, and Cnossen (2020), who proposed a goal competition model in PRIMs, as arising from memory dynamics in ACT-R, to account for the interaction of motivation and mental fatigue. Our model integrates the memory retrieval mechanism and Reinforcement Learning mechanism, assuming that the goal module assigns value to chunks representing goals, with this value representing the subjective reward associated with accomplishing the goal. Unlike ACT-R, PRIMs removes productions, representing procedural knowledge in the same way as declarative memory, in the form of operators and activation value associated with it. In their model, task motivation decreases naturally as a function of time, resulting in decreased task performance. The motivation of the task is quantified as the activation value of the goal. Different goals compete with each other, and the one with higher activation will be preferred. Interestingly, in

Herlambang et al.'s theoretical framework, motivation, or cognitive control allocation is represented as the process of multiple goals competing with each other, which is determined by a mechanism similar to memory retrieval in ACT-R. This novel framework first allows us to think about motivation in terms of memory.

In the remainder of this paper, we compare our motivation model to the EVC model of effort allocation and motivation proposed by Shenhav et al. (2013). We argue that ACT-R's procedural system provides an equivalent way of calculating the expected value of control as proposed in the EVC model. To prove that, we develop a simple effort allocation model in ACT-R, showing that ACT-R is capable of selecting the optimal strategy by weighing costs and rewards when making a decision, in line with EVC Theory. Further, we extend this simple model to a more complicated and realistic computational model of a cognitive interference task (the Simon task), augmenting it with the new motivation component. The result demonstrates that the proposed framework is compatible with the EVC model, and it helps us understand why cognitive systems vary widely in making decisions for engaging in effortful activities. Moreover, we propose a modeling approach for future ACT-R modelers that incorporates costs, rewards, and motivational components into cognitive function. All of the model and simulation codes and data are freely available at https://github.com/UWCCDL/ACTR-Motivation

3.3 COMPUTATIONAL MODELS

Motivation and Effort Allocation in an Abstract Model. To demonstrate the relationship between EVC theory and the proposed ACT-R motivational framework, we first present a simple, abstract ACT-R model and simulate the expected value of control predicted by the EVC model. To translate the continuous effort allocated in the EVC model, the abstract model assumes that different amounts of effort correspond to ten possible productions, indicated as

$P_1, P_2 \dots P_{10}$. The pre-conditions of these 10 productions are the same to guarantee that they are competing with each other. When the model starts running, only one of these 10 productions is selected based on the highest utility. Following this, an END production delivers a certain amount of reward at the end.

The 10 productions represent various ways to perform the task, which are associated with the calculation of expected payoffs and costs leading to optimal cognitive effort allocation. The cost of each production is represented by the time it takes to execute, which is controlled by a production-specific :AT parameter (for "Action time") in ACT-R. This parameter represents the effort associated with each production and, in the EVC framework, the associated cost of cognitive control. By default, it takes 0.05 seconds to fire a production, in this simple model, we assign different :AT to 10 productions in ascending order (0.01-0.1). Larger :AT suggests that the model needs to allocate more resources (time) in order to achieve the goal, while smaller :AT suggests that it could quickly finish the task, without spending more time on it. Specifically, production P_1 is assigned to the smallest :AT, and P_{10} is assigned to the largest :AT. Specifically, costs were allocated using the formula:

$$\text{Cost} = \frac{e^{ax}}{a} \qquad (3.3)$$

where a is a scale parameter set to 50 as the default constant.

To model expected payoffs, we set various amounts of rewards for 10 productions, in ascending order (0 - 10). P_1 is assigned to the lowest reward, while P_{10} is assigned to the highest reward. Following Musslick et al (2015)'s suggestion, we varied the costs of the different productions according to an exponential function and varied each production's probability of receiving a reward as a sigmoid function. Thus, assigned cost increases from P_1 to P_{10} exponentially, and the delivered rewards increase from P_1 to P_{10} following the sigmoid function. Simply put, P_1 spends the least cognitive resources but also has the lowest payoff, while P_{10} spends the most cognitive resources but has the highest payoff. Specifically, payoffs were allocated using the following equation:

$$\text{Payoff} = \frac{l}{1+e^{-k(x-x_0)}} \qquad (3.4)$$

where $l = 10$, $k = 1$, $x_0 = 5$ as the default constant.

Two experimental conditions were simulated, corresponding to the two theoretical cases discussed by Shenhav et al. (2006). The first is the effect of increased task difficulty. This was simulated by decreasing each production's probability of obtaining the reward. The second was the effect of an increase in the payoff. This was done by increasing the absolute value of M and, therefore, of the reward associated with accomplishing the task. In our framework, this is equivalent to simulating higher intrinsic motivation. We simulated the model 100 times per parameter set and 100 seconds (in ACT-R time) for each trial. During each trial, we recorded the counts of each selected production to estimate the probability of selecting a production. For each selected production, the received reward was also recorded to estimate the payoff.

It was expected that the probability of a production being selected would show the same pattern predicted by the EVC model. Specifically, we hypothesized that, under different combinations of rewards and costs, the model would assign the greatest utility (and, therefore, the greatest probability of being selected) to the production that maximizes the difference between rewards and costs. Both low-cost low-payoff productions (P_1, P_2), and high-cost high-payoff productions (P_9, P_{10}) are less likely to be selected than optimal cost-reward balanced productions (P_6, P_7, P_8).

Motivation and Effort Allocation in a Realistic Task. The simple ACT-R model of effort allocation described above seems to provide an adequate and mechanistic implementation of the EVC framework. This case, however, was highly stylized: the ten productions do not represent specific cognitive operations and their costs do not realistically reflect cognitive times; this level of detail is, by contrast, the very strength of ACT-R. To examine whether the motivation framework outlined above could be translated into a realistic ACT-R model of a cognitive task of effort we applied it to Stocco et al.'s (2017) model of the Simon task (Figure 3.1). The model was chosen because it is freely available (at github.com/UWCCDL/PSS_Simon), and includes a variety of interacting cognitive mechanisms (including memory retrieval, attention, and reward learning), thus providing a realistic testbed for a cognitive architecture. Furthermore, the Simon task is the same task used in an influential paper by Boksem et al. (2006) to study motivation. The Simon task requires participants to respond to visual stimuli by pressing a left button to one shape (e.g., a circle) and a right button to another (e.g., a square). Congruent trials are where the stimulus is displayed on the same side as the rule dictates (e.g. if participants see a circle, they should press the left button; if participants see a square, they should press the right buttle), while incongruent trials are on the opposite side. This paradigm was widely used in neuropsychological studies to assess the ability to inhibit cognitive interference that occurs when the processing of a particular visual property hinders the simultaneous processing of a second stimulus property.

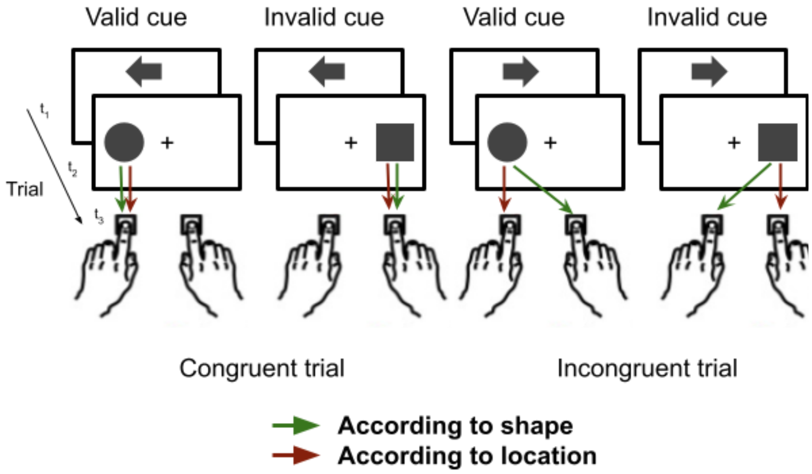

Fig 3.1. The demonstration of Simon Task.

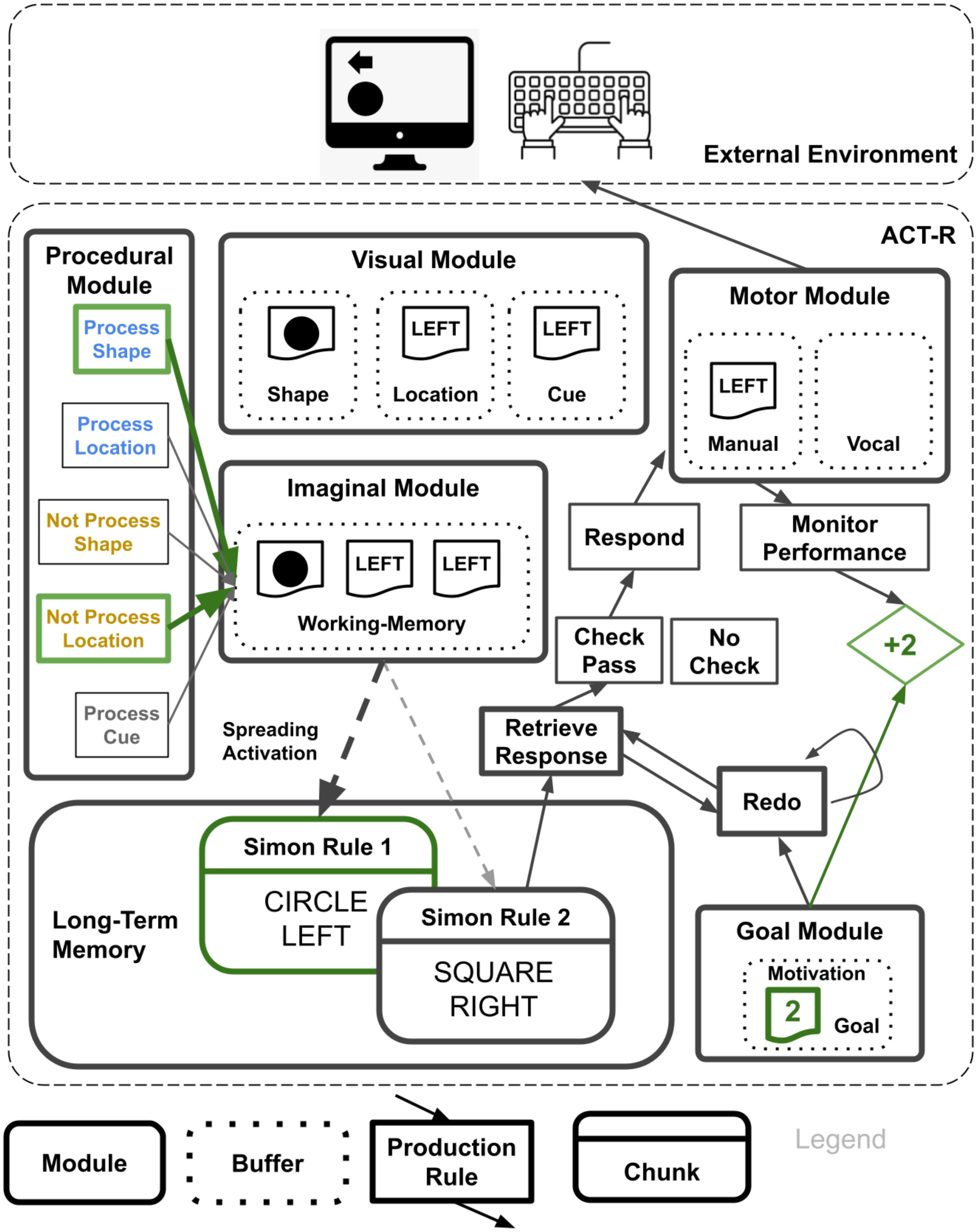

Fig 3.2. The motivation model of the Simon task in ACT-R

Figure 3.2 provides a complete overview of the model. It is composed of 4 main steps: (1) Encoding visual stimulus; (2) Retrieving a Simon rule; (3) Responding; and (4) Monitoring performance. The model starts by encoding a cue stimulus, and then it selects which dimension of the Simon stimulus to attend to: color or shape. After the stimulus is processed, the model retrieves the corresponding rule (e.g. press left if see a circle; or press right if see a square). The attended dimension provides spreading activation that facilitates the retrieval of the associated rule (a feature common to other response interference models in ACT-R: Lovett 2001; van Rijn 2009). The equation below (Eq 3.5) describes the activation of chunks calculated with a base-level learning function (B_i), which reflects the recency of previous retrievals, as well as a spreading activation component that reflects the degree to which a chunk matches the contextual components, i.e., the values of every slot j in every buffer k (See Chapter 2.3 Eq 2.4, Eq 2.5).

$$A_i = B_i + \sum_k \sum_j W_{kj} S_{ji} + \varepsilon \qquad (3.5)$$

Before a response is made, a "check" production executes a final verification step and, if it finds that the response is incorrect, attempts to re-allocate attention and retrieve another answer. The model contains additional assumptions (about the nature of competing productions that represent the contributions of different basal ganglia pathways in attention) that are not relevant to the goals of this study and will not be discussed. What is relevant, instead, is the nature of the "check" process. In the original paper, it was constrained to occur only once. In our extended simulation, we removed this limitation and allowed the model to check as many times as possible. The number of checks provides a natural way to model a decision threshold (akin to a decision boundary in Drift Diffusion Models) in ACT-R: repeated checks ensure greater decision accuracy at the price of greater response times. Because the number of checks performed also corresponds to how much effort is used to control attention, it provides a natural way to model the cognitive control effort in the task.

In addition, we incorporated a motivation value *M* in the Goal chunk and added a self-monitor production that assesses whether the response was correct and, if so, triggers a reward of magnitude *M*. Like the reward and cost parameter in ACT-R, *M* is also in time units, representing how many seconds the model is willing to continue working on the task. Note that the model will continue checking only if it finds the current response incorrect. Thus, if *M* is set high, the model would have more opportunities to correct its response. On the contrary, if *M* is small, it would have fewer opportunities to refocus attention.

To increase the task difficulty, Boksem et al. (2006)'s paradigm added cues stimulus, where 80% of the cues were valid. They identified an interesting post-error slowing effect in which participants tended to respond more carefully and slowly after they thought they had made a mistake. This process is believed to reflect adjustments in the allocation of cognitive effort, which is key to the EVC and our motivation framework. Critically, we verified that post-error slowing is not produced by Stocco's original model; thus, any success in reproducing this effect must be due to our additional changes.

In our model, fatigue is simulated by allowing the cost parameter (:AT) to exponentially increase over time (T) as outlined in Eq 5.6. The constants a and b were selected and set at 0.005 to ensure an optimal fit within the model's established parameters, and c represents the initial cost, by default as 0.05. For the purpose of simplification, T is treated as the trial number, but it could readily be converted to ACT-R time by picking appropriate constants a and b in future work. It's worth noting that we chose not to fit parameters to individual subjects in this iteration of our research, but this refinement could certainly be introduced in future analyses to further enhance the model's accuracy and applicability.

$$Cost(t) \quad = \; c \; - \; a(1 \; - \; e^{bt}) \qquad (5.6)$$

It was predicted that this model would be able to change strategies based on the probability of gaining rewards and costs. If it never checks, the likelihood of gaining rewards will become low because of many errors. If the model checks a lot, the expected payoff will also decline because of the increasing costs. Therefore, the model should weigh costs and rewards to decide the attempts of checking optimally.

We varied the *Motivation* parameter *M*, the task difficulty parameter VC (which represents the percentage of cues that are valid cues) as well as the initial cognitive control costs through the action time (:AT) parameter in ACT-R, which determines the time (and, thus the effort) needed to execute each production. The parameter space is shown in Table 3.1 below.

Table 3.1 *Model Parameter*

Parameter	**Value**	**Meaning**
M	0.5 - 10	Motivation
VC	0 - 1	Task Difficulty
AT	0.01 - 0.1	Cost of Control at T_0

3.4 RESULTS

Figure 3.3. (A) - (B), adapted from Shenhav et al. (2013), illustrates the changes in the relationship between payoff, costs, and control in the EVC model in both conditions. Figure 3.3

(C) - (D) demonstrates the relationship between cost, reward, and expected value of control in the abstract ACT-R model. As expected, our simple cognitive effort allocation model generated identical patterns of cognitive control allocation as the EVC model does. It selected the optimal production by weighing costs and rewards through utility learning in Reinforcement Learning. At a medium level of difficulty and a medium level of payoff (Figure 3.3 C), ACT-R selects the P_7, production most frequently because the utility of P_7 is the highest after subtracting costs from payoffs. As the task difficulty increases, ACT-R moves to select production P_9 most frequently. In terms of the EVC model, ACT-R now allocates more resources (a more costly production) to obtain rewards. If, on the other hand, the task difficulty decreases, ACT-R switches to a less effortful production rule (P_6), which guarantees similar rewards but with fewer costs (shorter times).

We observe similar patterns when the Payoff is manipulated (Figure 3.3 D). In lower payoff, ACT-R chooses P_4 most frequently. As payoff increases, ACT-R tends to allocate more resources to gain more rewards by selecting the higher-cost higher-reward production P_5.

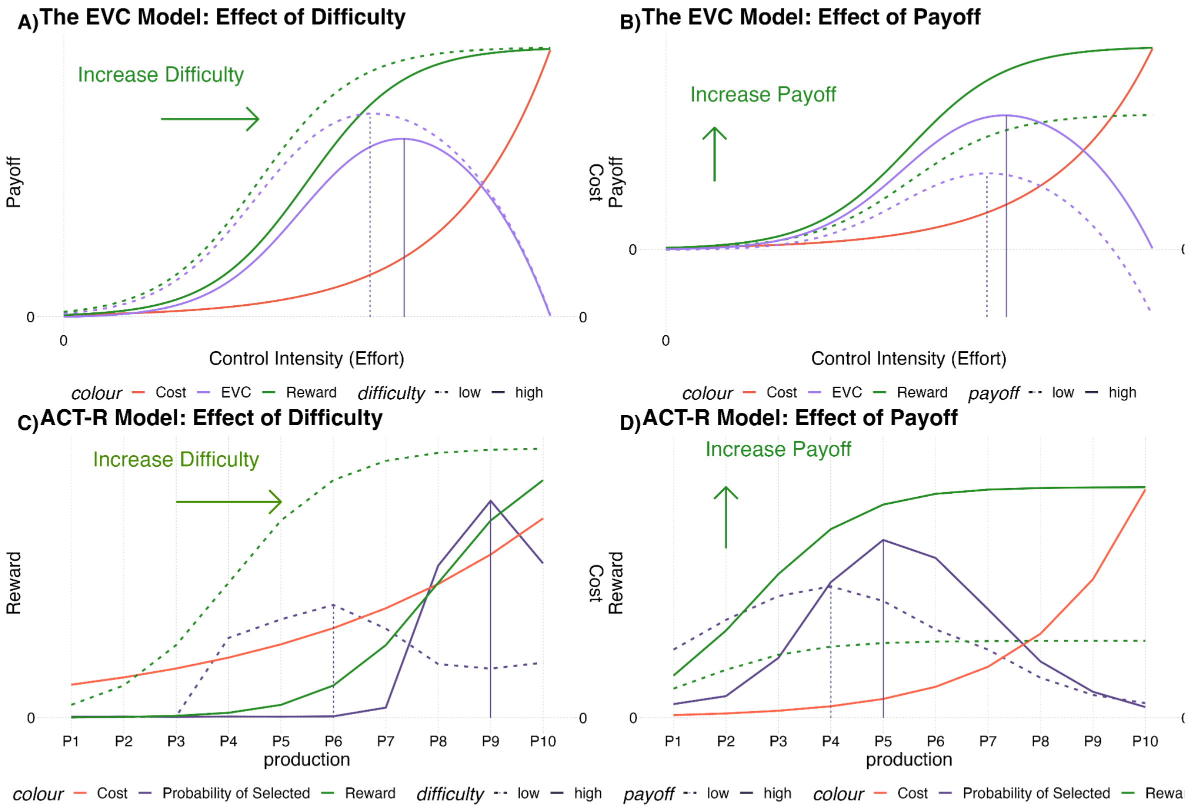

Fig 3.3. The Expected Value of Control Simulation Results. *The recreated plot of the EVC Model (A, B); the probability of production in the simple ACT-R model (C, D). In A) and B), the x-axis represents control intensity, red curves describe the costs increasing exponentially as control intensity, and green curves describe the payoffs curve. The purple curve describes the expected value of control. In C) and D), the x-axis represents 10 productions assigned with increasing costs and increasing rewards; Green dots denote the reward each production received; Red dots denote the cost (:AT) of each production; Purple dots denote the probability of each production being selected given the reward and cost parameter. Line types denote two conditions: increase the difficulty (A, C) and increase payoff (B, D). In (A) and (C), dashed curves describe the lowest task difficulty, while solid curves describe the highest task difficulty. In (B) and (D), dashed curves describe the lowest payoff (received rewards), and the solid curves describe the highest payoff (received reward).*

To test the validity of our model, we first compared simulated data to Boksem et al.'s (2006) empirical data. In the model simulations, we used a set of default parameters for accuracy and consistency. The ACT-R parameters were set as follows: *s*=0.2, *le*=0.63, *F*=1.0, *mas*=4.0, *d*=0.1, *egs*=0.2, *alpha*=0.4, and *imaginal-activation* = 3.0. Additional parameters included *dat*=0.04, *M*=1, *valid_cue_percentage*=0.8 for each of the single sessions. Each set of conditions was simulated 100 times using random seeds to ensure robustness and reliability of the model outcomes. Figure 3.4. confirmed that our model is able to reproduce the main Simon effects. Incongruent trials were associated with lower accuracy and longer response time than congruent trials; the same pattern was observed for invalid trials.

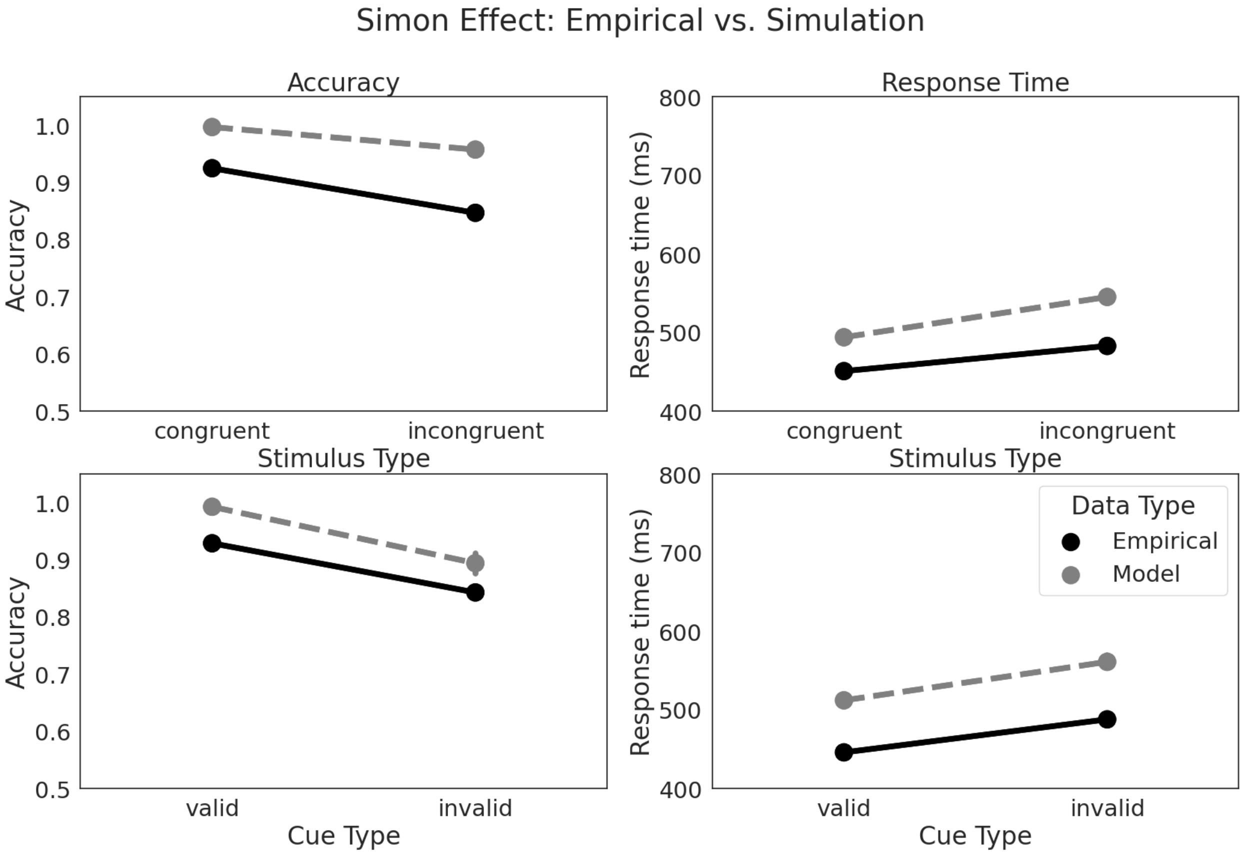

Fig 3.4. Model Simulation results of Simon effect vs. empirical findings of Simon effects (Boksem et al., 2006). *Black solid lines denote the simon effects from empirical data and gray dashed lines denote the simon effects simulated by our model. The default parameters used in model simulation are: s=0.2, le=0.63, F=1.0, mas=4.0, d=0.1, egs=0.2, alpha=0.4, imaginal-activation=3.0, dat=0.04, M=1, valid_cue_percentage=0.8, n_trials = 72, session =1, simulating for 100 times with random seeds.*

Interestingly, our model could naturally reproduce the post-error slowing effect observed in the empirical data (Figure 3.5). We utilized the same parameter set for consistency, but made modifications to align better with the paradigm established by Boksem et al. (2006). We ran 6 consecutive intervals 100 times and calculated the average response time following both correct and error trials for a more comprehensive comparison. Post-error slowing occurs when people tend to respond slower if errors are detected. Gehring and Fencski (2001) pointed out that this effect could be regarded as a componentry control mechanism that allows people to adjust cognitive control in order to improve consequent performance. Botvinick et al., (2001) tended to explain this phenomenon using the conflict monitoring account, arguing that the control strategy is adapted in a way that the cognitive control threshold is increased to deal with the incoming loss due to possible mistakes. Note that this effect could not be reproduced by the original model (Stocco et al., 2017) under any combination of parameters; in fact, we are not aware of any existing ACT-R model that accounts for post-error slowing. Thus, it is a unique consequence of the added motivation mechanism. Specifically, post-error slowing is a consequence of the model adjusting control after a mistake is made.

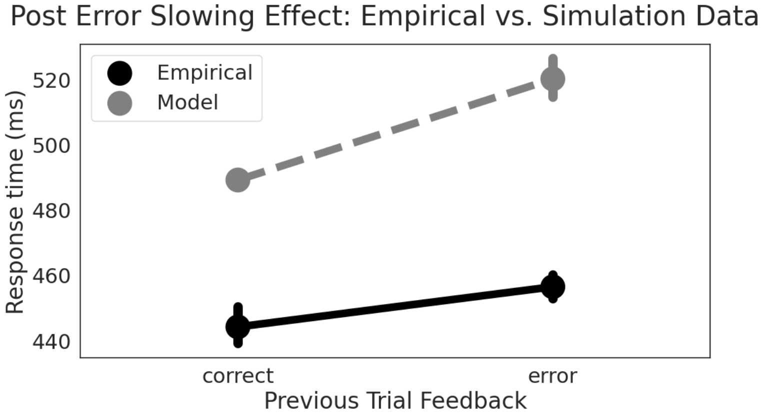

Fig 3.5. The post-error slowing effects from both empirical data (Bokesem et al., 2006) and model simulation data. *Response time (ms) is averaged on all intervals, where black solid lines denote the effects of empirical data and gray dashed lines denote the model simulation data. The default parameters used in model simulation are: s=0.2, le=0.63, F=1.0, mas=4.0, d=0.1, egs=0.2, alpha=0.4, imaginal-activation=3.0, dat=0.04, M=1, valid_cue_percentage=0.8, n_trials = 72, session =7, simulating 7 consecutive intervals for 100 times with random seeds.The standard error for both empirical data and model data is shown in the plot.*

In the Simon model, the degree of cognitive control is determined by how often the CHECK production is employed before a response is made. Additional firings of the CHECK productions result in repeated allocations of attention and, thus, more time spent before making a response. As hypothesized, we found that a model with a lower M ($M < 2.5$) would check only once or never check, while a model with a high M ($M > 7.5$) tends to check more. For example, when $M < 2.5$, the model performs an average of 0.54 checks, when $M < 7.5$, the model performs 0.81 checks, and when $M >= 7.5$, it performs an average of 1.01 checks.

As predicted by EVC theory, the relationship between motivation, number of checks, and utility of the CHECK production are complex and nonlinear. To examine this relationship, we fixed the parameter VC to 0.5. Figure 3.6. represents the resulting relationship between the costs, rewards, and allocated control. In the figure, the x-axis represents the intensity of control as the number of firings for the CHECK production, and the y-axis represents rewards, costs, and utilities in time units. The cost curve (red line) is represented by the total response time the model takes as a function of the count of checking. Moreover, in our self-monitoring process, once the model verifies that the response was correct, a reward equal to *M* is delivered. The utility of the CHECK production (purple line) represents the expected value of control in the EVC model.

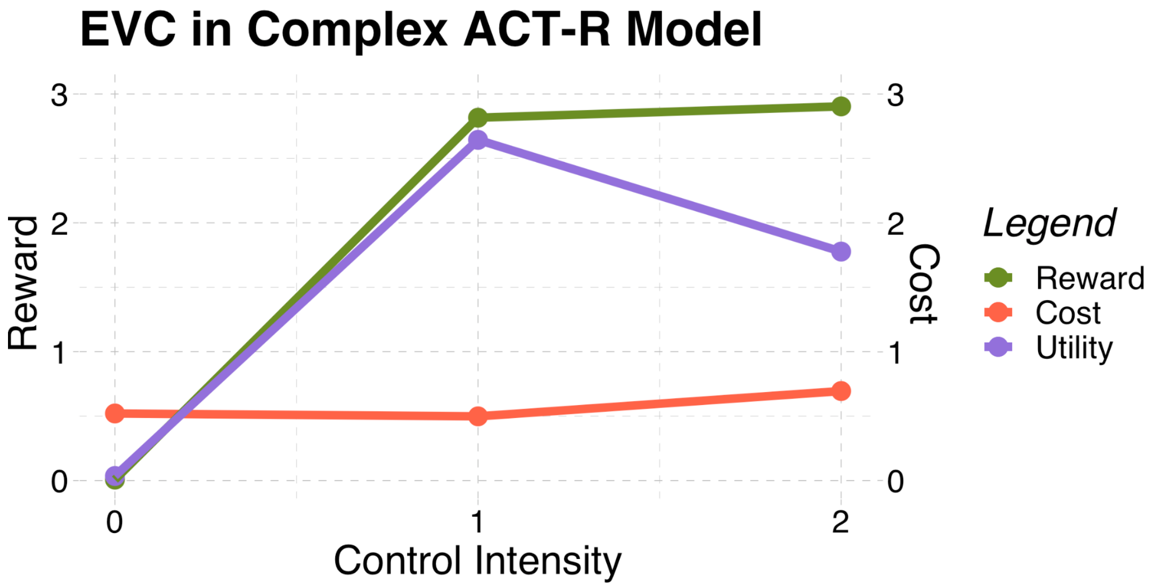

Fig 3.6. The Expected Value of Control as a Function of Control Intensity, Reward and Cost in the Simon Task Model. *Control intensity is expressed as the number of firings of the CHECK production. Note that, even when higher rewards would be possible at a higher level of control, the model naturally shifts to the amount of control that maximizes the difference between rewards and costs.*

In line with the EVC model, our ACT-R model predicts that the model will be encouraged to invest more effort if expecting a higher payoff, but as costs increase, the expected reward decreases, and the model decides to stop allocating more cognitive control by reserving resources. Note that, although the model could achieve greater performance through greater control, it naturally sets to an estimated value of one check per trial because, at this level, the payoff is maximal: additional checks have many diminishing returns. Incidentally, this is precisely the number of checks that were determined to yield optimal results in Stocco et al. (2017) and Lovett (2005).

Interestingly enough, our model captures the impact of mental fatigue on cognitive task performance. Figure 3.7 presents a compelling comparison between the empirical data collected by Boksem et al. (2006) and the results derived from our model's simulations. This comparison clearly illustrates the impact of progressive fatigue on cognitive task performance. Across 1-6 intervals, representing increasing levels of fatigue, both sets of data show a notable decline in performance. This deterioration is evidenced by a gradual slowing in response times and a decrease in accuracy. These trends underline the detrimental effects of fatigue on cognitive function and demonstrate how effectively our model replicates these effects in a controlled, simulated environment.

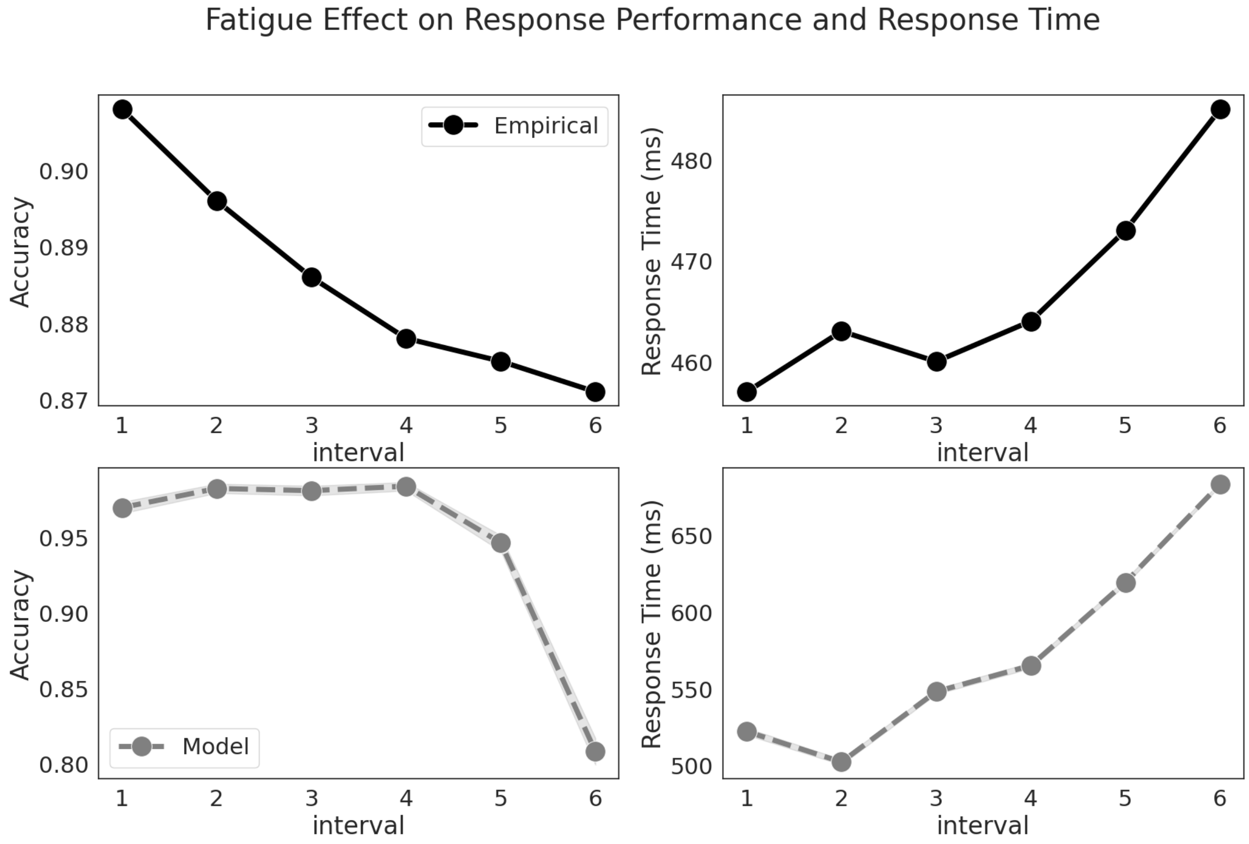
Fatigue Effect on Response Performance and Response Time

Fig 3.7. Fatigue effect on performance from both empirical data (Bokesem et al., 2006) and model simulation. *Interval 1-6 indicates 6 successive sessions of the task, suggesting the level of fatigue introduced over time. The default parameters used in model simulation are: s=0.2, le=0.63, F=1.0, mas=4.0, d=0.1, egs=0.2, alpha=0.4, imaginal-activation=3.0, dat=0.04, init_cost=0.04, M=1, valid_cue_percentage=0.8, session = 6, simulating for 10 times with random seeds.*

3.5 DISCUSSION

In this paper, we have proposed a mechanistic interpretation of motivation within the ACT-R cognitive architecture. Specifically, we propose that motivation can be modeled by assigning a value *M* to the current model's goal and translating this value as the reward R_t that is triggered when the goal is accomplished. With this mechanism in place, ACT-R's utility learning mechanism then provides a way to adjust the specific combination of productions that are used to perform a task. Because in ACT-R, rewards and time spent on a task are expressed on the same scale (and rewards are adjusted by the time elapsed), the motivation parameter M can be equivalently expressed as the subjective reward associated with accomplishing the goal and the maximum amount of time that the model is willing to spend on the task. We first demonstrated, using a simple abstract model, that this mechanism is equivalent to EVC theory. We then showed how this mechanism can be easily implemented in an existing model of a common laboratory task (the Simon task) and used to account for experimental effects that would otherwise go unmodeled, such as post-error slowing, the effect of difficulty, and even fatigue. All of these effects can be understood as ways in which the model flexibly copes with changes in task demands.

A number of limitations must be acknowledged. First, the level control intensity is quantified by the counts of checking attempts as a discrete variable. By using only counts of checking attempts, this model may miss the finer variations and continuity in participants' cognitive efforts and control strategies during the task. Employing more continuous measures or exploring alternative approaches would allow for a more detailed and nuanced examination of the cognitive processes involved, enhancing the overall insights gained from the study. Future work will be needed to address these issues and expanding our model to represent the control intensity with a continuous variable could be the next step of research. Moreover, individual variability in motivation could be examined in future modeling work, specifically how motivation affects the response time rather than accuracy for individuals putting different priorities in speed vs. accuracy tradeoffs (Boksem et al., 2006).

These limitations notwithstanding, we believe that our results are noteworthy for three reasons. In addition to providing a way to implement motivation into ACT-R, our framework provides a more complete view of the role of the Goal module in ACT-R. In the ACT-R architecture, despite having different neural underpinnings, the functions of the Goal module and Imaginal module are practically indistinguishable. Both modules serve to represent the status of the task and aid in its progression. By establishing a connection between the Goal module and the amount of reward generated, using a parameter *M*, this framework offers an interpretation that aligns better with neuroscientific data. It also provides a connection to the original interpretation of the goal in previous versions of ACT-R, as well as the original production selection mechanisms. Finally, it provides a way to better-fit models at the individual levels, decoupling the effects of individual capacity and motivation.

Chapter 4. Study 2: Predicting Individual Strategy Selection From Neural Connectivity

4.1 Abstract

Experiential decision-making can be explained as a result of either memory-based or reinforcement-based processes. Here, for the first time, we show that individual preferences between a memory-based and a reinforcement-based strategy, even when the two are functionally equivalent in terms of expected payoff, are adaptively shaped by individual differences in resting-state brain connectivity between the corresponding brain regions. Using computational cognitive models to identify which mechanism was most likely used by each participant, we found that individuals with comparatively stronger connectivity between memory regions prefer a memory-based strategy, while individuals with comparatively stronger

connectivity between sensorimotor and habit-formation regions preferentially rely on a reinforcement-based strategy. These results suggest that human decision-making is adaptive and sensitive to the neural costs associated with different strategies.

4.2 BACKGROUND

Two distinct frameworks have been proposed to explain how experience influences human behavior. According to one class of theories, previous decisions and their outcomes are stored as traces in episodic long-term memory, and decision-making is controlled by which memory is easier to retrieve (Gonzalez & Dutt, 2011; Stewart et al., 2006): we will refer to this as the *declarative* framework. The other class of theories suggests that decision-making based on experience is governed by the basic principles of reinforcement learning, in which the value of the outcome of each previous decision incrementally modifies an internal cached value associated with each option (Niv, 2009; Lee et al., 2012); when making a decision, the option with the highest expected value is chosen. We will refer to this as the *procedural* framework.

Despite making similar predictions about the outcome of decision-making processes (Chelian et al., 2015), these two mechanisms depend on different neural resources. In the *declarative* framework, decisions from experience depend on the brain circuits involved in the storage and retrieval of episodes in long-term memory, such as the medial temporal lobe (Knowlton et al., 1994) the lateral prefrontal cortex (Badre & Wagner, 2007), and the medial frontal and parietal regions associated with the default mode network (Raichle & Snyder, 2007; Sestieri et al., 2011). In the *procedural* framework, decisions from experience rely on brain circuits for implicit reward learning and habit formation, such as the basal ganglia (Foerde et al., 2006), the supplementary motor area (Holroyd et al., 2004), the cerebellum (Taylor & Ivry, 2014), and the specific sensorimotor regions involved in the task itself (Hill & Schneider, 2006).

Both systems are concurrently active at any given time (Foerde et al., 2006; Poldrack et al., 2001), so why should individuals choose to rely on one system over the other? One likely explanation is that optimal behavior in humans is inherently limited by processing capacity (Adams & Simon, 1962), therefore individuals must adopt different procedures to maximize the outcome of a decision while consuming the minimum amount of cognitive resources (Payne et al., 1993; Payne et al., 1988; Gigerenzer, 2008). Because the cost of a decision also depends on the neural resources it consumes, individuals with different neural characteristics would make decisions in different ways. This paper puts forward the hypothesis that individuals rely on *declarative* or *procedural* processes based on the relative neural efficiency of their corresponding brain circuits.

A common indicator of neural efficiency within a circuit is their functional resting-state connectivity, that is, the degree of correlation between the spontaneous neural activities of different regions in that circuit (Shen et al., 2017). Higher correlations at rest reflect tighter coupling of neural dynamics and greater exchange and integration of communication between regions (van den Heuvel & Hulshoff Pol, 2010). Thus, relatively higher functional connectivity within the *declarative* or *procedural* circuit should predict greater reliance on the corresponding system when making decisions.

An important obstacle in testing our hypothesis is that the choice between the two systems does not depend solely on their mental and neural costs but also on their relative effectiveness in a given task. For example, it has been argued that the *procedural* system would be preferred when decisions are probabilistic and the stimuli are difficult to verbalize (Knowlton et al., 1994; Poldrack et al., 2001; Frank et al., 2004). If, for some reason, one system is better suited for a given task than the other, an individual preference based on neural efficiency would be overridden by the effectiveness of the alternate system. Investigating the relationship between brain connectivity and preferred decision-making processes requires a task where the *procedural* system is equally as effective as the *declarative* system strategy. One such task is the Incentive Processing Task (IPT) (Delgado et al., 2000), which guarantees that either system yields the same expected payoff. In this task, participants repeatedly guess whether a hidden number is greater or smaller than five by pressing one of two buttons, and receive monetary feedback for correct guesses. Once the choice is made, the number is revealed, and feedback is provided (Figure 4.1A). Unbeknownst to participants, the number is chosen after their decision is made and follows a predefined feedback schedule. Under these conditions, preferences for either the *declarative* or *procedural* decision-making systems should depend only on the neural costs for each individual participant. Because the number of wins and losses is fixed and predefined, behavioral data from the IPT cannot be analyzed using accuracy measures. Instead, individual behavioral differences can be measured by computing the probability of choosing a different option (i.e., from “win” to “lose” or vice versa) after receiving feedback from the previous trial, referred to as the shift probability. Shift probabilities were computed for the different types of feedback (“win,” “lose” or “neutral” when the number is exactly five) and the two different types of experimental blocks that were used in the task (“Mostly Lose," in which 6 loss trials were pseudo-randomly interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials, and “Mostly Win”, in which 6 reward trials were pseudo-randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials). Under these conditions, individual preferences for *declarative* vs. *procedural*-based decision-making should maximally reflect the underlying efficiency of the corresponding circuit.

To test this hypothesis, we analyzed a subset of 200 participants from the Human Connectome Project (Van Essen et al., 2013) for whom neural and behavioral measures from the IPT as well as resting-state fMRI data were available. Figure 4.1B illustrates the methodology of this study. To determine whether a participant relied on *procedural* or *declarative* mechanisms, each participant's behavior was fitted to two parametrized computational models, one implementing *declarative* decision-making and one using the *procedural* system.

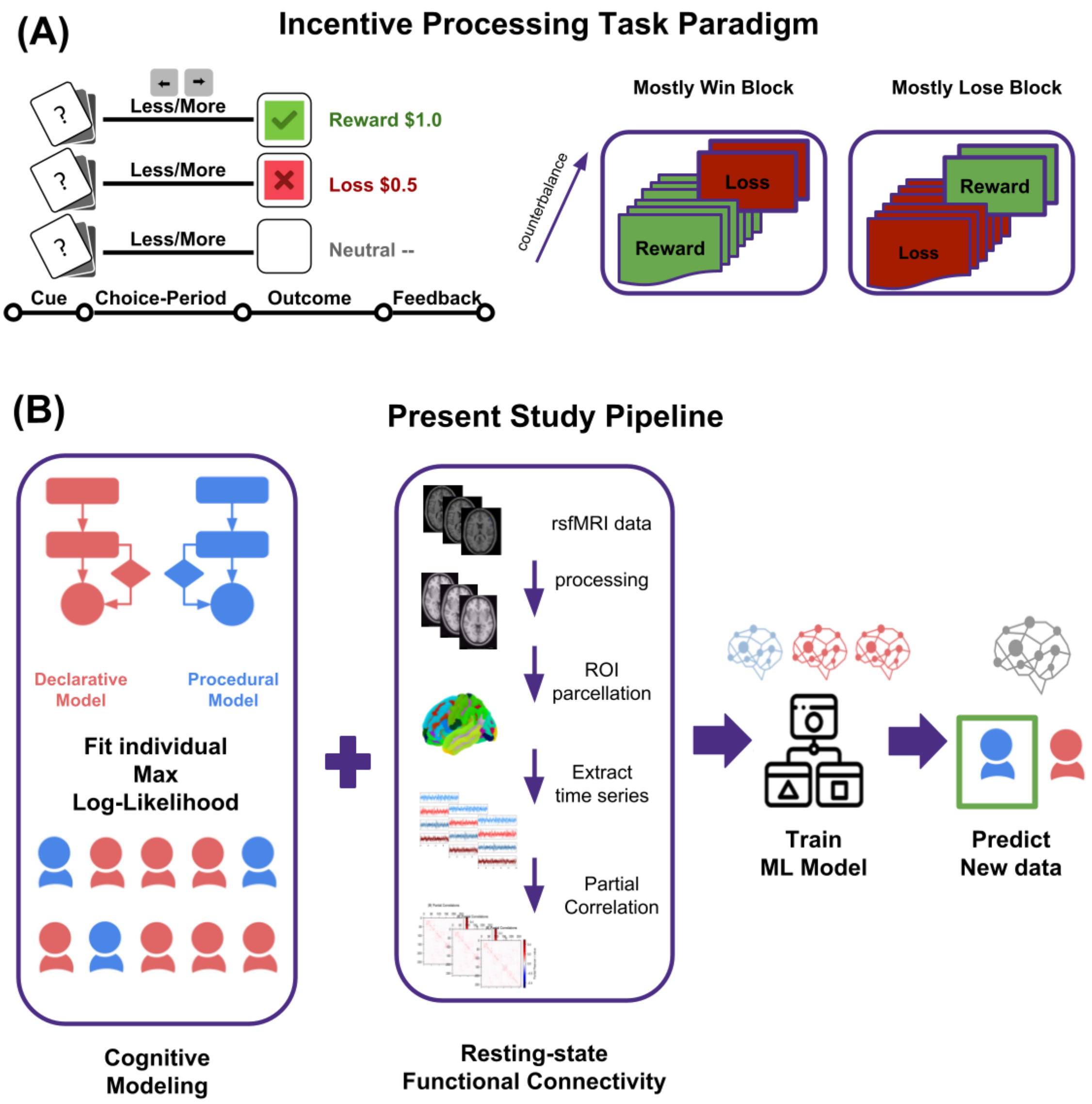

***Fig 4.1. (A) Incentive processing task paradigm. (B) The pipeline of the present study.** First, each subject's behavioral data are fit by two distinct cognitive models, declarative vs. procedural model through maximum log-likelihood. Second, each subject's neuroimaging data are processed to construct resting-state functional connectivity matrices. Lastly, machine learning models are trained to predict an unseen subject's decision-making process from the neuroimaging data.*

The episodic control model is based on the memory model of Anderson and Schooler (Anderson & Schooler, 1991) that is included in the ACT-R architecture (see Chapter 2), and follows the multiple trace theory (Nadel et al., 2000). In this model, the outcome of every decision is stored as an episodic memory. After encoding a memory, without any intervention, that memory starts to be forgotten, and the availability of each memory trace decays according to a power law (Newell & Rosenbloom, 1982). Multiple experiences of the same decision-outcome pair, however, result in the accumulation of traces of the same memory, increasing the *activation* of that memory. As noted in Chapter 2 (See Chapter 2.3, Eq. 2.4 Eq 2.5), the activation of a memory m at time t reflects the balance between these two effects, and is given by below equation:

$$A_t(m) = \sum_n \log (t - t_n)^{-d} \tag{4.1}$$

In this equation, t_n is the time at which the same decision/outcome pair has been experienced for the n-th time and d is an individual-specific forgetting rate (Zhou et al., 2021; Sense et al., 2016). A decision is guided by the most active memory of past interactions. If the retrieved memory is of a "win" outcome, the model repeats its associated decision, (e.g., choosing "more" if the "win" was associated with a "more" choice). If the memory is of an associated loss, the model executes the *opposite* action (e.g., choosing "less" if the loss was associated with a "more" choice). The probability that a given memory m will be retrieved is controlled by a Boltzmann function with noise temperature v (See Chapter 2.3 Eq 2.6):

$$P(m) = e^{A(m)/v} / \sum_i e^{A(i)/v} \tag{4.2}$$

The *procedural* model is based on the reinforcement learning framework (Niv, 2009). In this model, the two possible decisions ("more" or "less") are represented as two actions associated with the cue (in this paradigm, this is the presentation of the two cards). As reviewed in Chapter 2.3 (see Eq 2.8), each action *a* has an associated value $V(a)$ that is updated over time based on the difference between the expected and the actual outcome (Rescorla & Wagner, 1972). For example, receiving a "win" feedback after making a "more" choice will increase the value of the "more" action, $V(more)$, making it more likely to be selected in future choices, while receiving "loss" feedback will decrease the value of the "more" action, making it less likely to be selected in future choices. Formally, after the *n*-th decision, the value $V_n(a)$ of an action is updated using the temporal difference method (Sutton, 1988):

$$V_n(a) = V_{n-1}(a) + \alpha[R - V_{n-1}(a)] \tag{4.3}$$

In Equation 3, R is the reward associated with an outcome, (1 for "win", -0.5 for "lose", corresponding to the dollar amount won by participants), and α is the learning rate. The probability that an action *a* will be selected is controlled by a Boltzmann function with noise temperature τ:

$$P(a) = e^{V(a)/\tau} / \sum_i e^{V(i)/\tau} \tag{4.4}$$

Thus, each model depends on one learning parameter (*d* or α) and one noise parameter (ν or τ).

Both models were fit to each participant using maximum likelihood estimation (Myung, 2003), i.e., by fitting their parameters to maximize the probability of generating the specific series of decisions made by that participant (see Methods). Each participant was then assigned to either the *declarative* or *procedural* group based on which version of the model had the greatest likelihood of producing their observed behavioral data.

Although indistinguishable in most cases, the two models behave differently in the IPT. Because, in the *declarative* model, the availability of a memory grows with the number of times it has been experienced as the task continues, feedback from the most recent decision has less influence on future choices, which are instead affected by the longer history of previous decisions. The *procedural* model, on the other hand, does not maintain a memory of previous choices and adjusts the value of an action after every feedback; thus, it is more sensitive to the most recent outcome. Because of this, we expect that individuals best fitted by the *declarative* model would be less prone to changing their choice preferences immediately after feedback, and thus would exhibit smaller shift probabilities. And because the two mechanisms rely on different neural substrates, we expect that regional differences in BOLD activation would reflect the neural cost associated with each. Therefore, we expect individuals best fitted by *declarative* models to show greater activity in regions associated with episodic encoding, such as the hippocampus, and retrieval, such as the ventral prefrontal cortex and default mode network. Correspondingly, we expect individuals best fit by the *procedural* model to display greater activity in the circuits associated with habit formation, such as the basal ganglia, the cerebellum, and the visuo-motor regions involved in interacting with a computer task. Finally, we expect the type of decision-making processes used to depend on the patterns of functional connectivity at rest. To test this latter prediction, a classifier was trained to predict whether a participant would be best fit by a *declarative* or *procedural* model from the same individual's functional connectivity data. We expect that functional connections that successfully predict the reliance on the *declarative* mechanism for decision-making would be found between memory encoding and retrieval regions. Conversely, we expected that functional connections predictive of reliance on the *procedural* mechanism reward-based learning would be found in sensorimotor cortices and the basal ganglia circuit.

4.3 METHODS

Materials. This study analyzed both behavioral and neuroimaging data obtained from a subset of the Human Connectome Project (HCP) dataset (Van Essen et al., 2013). Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research and by the McDonnell Center for Systems Neuroscience at Washington University. A total of 199 participants (111 females, 85 males, and 3 did not disclose) who completed both sessions of the task-based fMRI gambling game were included in this study and fit by the two ACT-R models. By excluding 23 participants who missed resting-state fMRI scanning, a total of 176 participants were fit by a predictive LASSO model. All participants were healthy adults with no neurodevelopmental or neuropsychiatric disorders. The experimental protocol, subject recruitment procedures, and consent to share de-identified information were approved by the Institutional Review Board at Washington University.

<u>*The Incentive Processing Task*</u>. This incentive decision-making task was adapted from the gambling paradigm developed by Delgado and colleagues (2000). Participants were asked to guess if the number on a mystery card (represented by a "?", and ranging from 1 to 9) was more or less than 5. After making a guess, participants were given feedback, which could take one of three forms, "Win" (a green up arrow and $1), "Loss" (a red down arrow and -$0.50), or "Neutral" (a gray double-headed arrow and the number 5). The feedback did not depend on the subject's response, but was determined in advance; the sequence of pre-defined feedback was identical for all participants. The task was presented in two runs, each of which contains 64 trials divided into eight blocks. Blocks could be "Mostly Lose" (6 loss trials pseudo-randomly interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials) or "Mostly Win" (6 win trials pseudo-randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials). In each of the two runs, there were two "Mostly Win" and two "Mostly Lose" blocks, interleaved with 4 fixation blocks (15 seconds each). All participants received money as compensation for completing the task, and the amount of reward is standard across subjects.

<u>*Analysis*</u>. This study employed the "minimally preprocessed" version of resting-state fMRI data and incentive processing task fMRI data, which has already undergone a minimal number of standard preprocessing steps including artifact removal, motion correction, normalization, and registration to the standard MNI ICBM152 template. Additional preprocessing steps were performed using the AFNI software (Cox, 1996; Cox, 2012), including despiking, spatial smoothing with an isotropic Gaussian 3D filter FWHM of 8 mm, and removal of linear components related to the six motion parameters and their first-order derivatives.

Functional connectivity measures were constructed from the HCP resting-state data using Power et al.'s whole-brain parcellation (Power et al., 2011). This parcellation was used to construct a 264 Region of Interest (ROI) functional atlas, with each ROI containing 81 voxels. This parcellation atlas is defined in the MNI space and was applied to all participants in the HCP dataset. The extraction of the time series and calculation of the connectivity matrices was performed using R (2020) and Python. Pearson correlation coefficients and partial correlation coefficients between the time series of each brain region were calculated for each participant, resulting in a 264 × 264 symmetric connectivity matrix for each session for each subject. The average correlation coefficients across subjects were calculated by first transforming each r value into a *Z*-value, and then retransforming the average *Z* value back into an equivalent *r* value using the hyperbolic tangent transformation (Silver & Dunlap, 1987).

For task-based fMRI data analysis, we specified the first-level analysis model and estimated the parameters corresponding to the difference between "Mostly Win" and "Mostly Lose" blocks, as in (Barch et al., 2013). The resulting contrast maps for each subject were then used in a second-level weighted t-test between declarative and procedural groups. The test was implemented using AFNI's 3dttest++ software, and its weights corresponded to the absolute difference in log-likelihood between the best-fitting declarative and best-fitting procedural model. This way, the contribution of each observation was proportionally scaled to the evidence favoring each participant's assignment to their groups. The statistical significance level was set at a significance level of $q < 0.01$ corrected for multiple comparisons using a False Discovery Rate (FDR, Benjamini & Hochberg, 1995) procedure.

4.4 COMPUTATIONAL MODELS

Episodic Memory Model. The episodic memory model relies on the declarative model of Anderson and Schooler (Anderson & Schooler, 1991), currently implemented as part of the influential ACT-R cognitive architecture (Anderson, 2009). When presented with a mystery card, the model retrieves a previous episodic memory in which the number was guessed successfully. After being presented with feedback, the model encodes a new memory associating the action with its outcome. Memories are retrieved based on their activation, a noisy quantity that depends on the frequency and recency with which the decision-outcome episodes have been experienced (Eq. 1).

Procedural Model. By contrast, the procedural model represents the possible actions of the decision-making processes as competing stimulus-response rules, and reinforcement learning is used to increase the use of the rule that leads to the best outcomes. Instead of encoding each trial as a memory of action and associated feedback, the model has two competing actions that implement the "More" and "Less" decisions. When presented with the mystery card, the model chooses one of the rules to execute based on its expected value. Initially, both rules have equal value, and one will be chosen at random. After each decision, the model is presented with a "Win", "Lose", or "Neutral" response, and this feedback is encoded as the reward term in the reinforcement learning equation (Eq. 2) (+1 for a "Win" result, -0.5 for a "Lose" result, and 0 for a "Neutral" result). Positive rewards will encourage the model to repeat the associated action, while a sequence of losses will decrease the value of an action and encourage the selection of the alternate action.

Model Fitting. Model parameters were fit to maximize their likelihoods given the data. The likelihood of a model M with parameters θ given a vector of data $\mathbf{x}$, indicated as $L(M, \theta \mid \mathbf{x})$, is the probability of observing the data, given the model and its parameter, that is $L(M, \theta \mid \mathbf{x}) = P(\mathbf{x} \mid M, \theta)$. The data **x** consists of a vector of decisions $x_1, x_2, \ldots, x_N$:

$$L(M, \theta \mid \mathbf{x}) = P(x_1 \mid M, \theta) \cdot P(x_2 \mid M, \theta) \cdot \ldots \cdot P(x_N \mid M, \theta) = \prod_i P(x_i \mid M, \theta)$$

Since the product of conditional probabilities becomes vanishingly small, it is common to use *log*-likelihood:

$$\log L(M, \theta \mid \mathbf{x}) = \log \prod_i P(x_i \mid M, \theta) = \sum_i \log P(x_i \mid M, \theta) \qquad (4.5)$$

Note that, because both models are *memory* models, the probability of decision x_i depends on the *history* of previous decisions $x_{i-1}, x_{i-2}, \ldots, x_1$ and their outcomes, i.e. $P(x_i \mid M, \theta) = P(x_i \mid M, \theta, x_{i-1}, x_{i-2}, \ldots, x_1)$. In both models, the history of previous decisions is implicitly represented in their respective memory representations, that is, the activations of episodic memories and the values of actions. Thus, the probabilities of generating a response x_i can be derived by Equations 2 and 4, provided that the history of previous decisions and outcomes is recorded in the corresponding activations and values.

To do so, each model is run forward, first calculating the probability of producing the participant's response in the current trials; next updating the cumulative log-likelihood; and finally adjusting the activation of episodic traces and the value of the decision based on Equations 1 and 3 before moving to the next trial. For each participant, the model simulates the exact series of 64 trials as they experienced them, in the exact order and with the same inter-trial and inter-block intervals.

In order to evaluate the goodness-of-fit for individual fitting, we estimated maximum Log-Likelihood across the parameter space, i.e. we selected the value of parameters θ^* such that $\theta^* = \text{argmax}_{\theta} L(M, \theta \mid \mathbf{x})$. Because no closed-form solutions exist to derive the maximum likelihood of Equations 1 and 3, the best-fitting parameter values were identified using Powell's optimization algorithm (Powell, 1964), as implemented in Python's SciPy package (Virtanen, Gommers, Oliphant, Haberland, Reddy, Cournapeau, Burovski, Peterson, Weckesser, Bright, Walt, et al., 2020). This method was chosen because it does not require explicit derivatives and allows us to specify meaningful bounds for parameter values. All parameters were constrained within the bounds [0, 2]; they are all meaningless when negative and lead to instability and essentially random choices when > 2.

The starting parameter values for the algorithm were $\alpha = 0.1$, $\tau = 0.1$ for the procedural model, and $d = 0.5$, $\nu = 0.1$ for the declarative model. The starting values of α and d were chosen as the most common values found in their respective modeling frameworks; they are also the default values of the corresponding parameters in ACT-R.

Supervised Classification Model. To explore if individuals' behavioral differences between declarative and procedural strategies are indicated by an individual's underlying brain structure, we trained three of the most commonly used supervised classification models (Logistic regression model, Decision Tree model, and Random Forest model), using resting-state functional connectivity as the input variable, and predicted the probability of a participant being labeled as either preferring the declarative or procedural strategy. Considering the equally high accuracy (> 0.8) among these classification models (Random Forest Model (accuracy = 0.92) slightly outperformed the other two models), we are confident to say that our Machine Learning models work very well in predicting the strategy selection from an individual's resting state neuroimaging data.

In addition to the predictive power of machine learning models, another very important dimension we need to carefully consider in Machine Learning related research is the model's interpretability. While some ML models are excellent at predicting outcome variables, as complexity grows exponentially (as in Deep neural networks), they become a black box that is fundamentally difficult to interpret. Therefore, choosing an appropriate ML model with reasonable predictability and interpretability is critical. We chose the logistic regression model because it is a simple and powerful binary classification model that has been widely used in many fields of research. It could be used to predict the likelihood of an event happening or a choice being made. Rather than fitting data as a straight line, the logistic regression model uses the logistic function to constrain a value between 0 and 1. Specifically, the output value x of a linear model is transformed into a corresponding probability $P(x)$ by the function:

$$P(x) = 1 / (1 + e^{-x}) \qquad (4.6)$$

In order to handle an imbalanced dataset with unequal target labels, upsampling was applied by randomly adding data from the minority class. We also applied individualized weights to each training sample, which is the absolute difference of maximum log-likelihood between two models. Specifically, for subjects who are better distinguished by either of the two models (declarative vs. procedural), we increased the weights of these data points in later ML model training, while for those who had a very close fit maximum log-likelihood between two models, the training procedure was less reliant on these samples. Having 69,696 (264 ROI × 264 ROI) connections, we want to select only the most important connections contributing to the prediction, therefore, LASSO regularization was applied to the Logistic Model. LASSO is a machine learning regression analysis technique that performs both variable selection and regularization in order to improve the prediction accuracy and interpretability of the computational model. It can reduce model complexity by penalizing large numbers of coefficients and also prevents overfitting which may result from a simple linear regression analysis. In LASSO, the best values associated with each regressor are chosen as to minimize a loss function that includes the ordinary squared errors as well as a penalty term for each regressor's β weight:

$$Loss = (y - \hat{y})^2 + \lambda \sum_{i}^{n} |\beta_i| \qquad (4.7)$$

where the tuning parameter λ controls the degree of penalty: for greater values of λ, more coefficients are forced to become 0.

To account for the large disparity between the number of participants and the number of predictors, we performed a grid search cross-validation using the *glmnet* package in R (Friedman et al., 2010) to determine the best value for the fit hyper-parameter $\lambda_{regular}$. To alleviate the potential problems of small sample size in neuroimaging studies, Vabalas and colleagues (Vabalas et al., 2019) used Nested CV approaches to produce robust and unbiased model performance regardless of sample size. Following their suggestions, we fit the model with *n*-iteration nested cross-validation (*n* = 200) to determine the optimal hyperparameter λ. For each iteration, the dataset was randomly split into training and testing (the ratio of training to testing is 1:4). Instead of adopting a computationally expensive Leave-One-Out (LOO) method, we adopted the *k*-folds cross-validation (*k* = 20) method. The whole dataset was randomly split into 20 folds and trained on 19 folds of samples. The prediction was made on the remaining one fold of samples to obtain the best hyperparameter value λ_k, which gave the lowest classification error. With each best lambda λ_k, the model was restrained and made predictions on testing datasets. Then the process was repeated k times. To guarantee maximum generalizability, we chose the median of lambdas (λ_{nested}). This approach allowed us to obtain a reliable and unbiased assessment of the LASSO model's performance while accounting for the variability in the data and the selection of hyperparameters.

The mean accuracy score, true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR) were calculated across all folds to evaluate the overall performance of the model, taking sample weights into account. By definition, the receiver operating characteristic curve (ROC) demonstrates the performance of a classification model by plotting the relationship between TPR vs. FPR at different classification thresholds. We calculated the AUC (Area under the curve), which is one of the most important metrics for evaluating a classification model’s performance; as the AUC of a model approaches 1, the model approximates an ideal, perfect classifier. It provides information about how well a classification model is capable of distinguishing between classes.

4.5 RESULTS

Reliability of Model Identification. Because our results depend on correctly identifying the model that best represents each participant, a model recovery procedure (Wilson & Collins, 2019) was run to ensure that our identification process was reliable. In this procedure, simulated participant data was generated by instantiating a model with random parameter values (uniformly drawn within their plausible boundaries; see Methods); the model-fitting procedure was then applied, treating the simulated data as participant data, and the accuracy to which the original model could be correctly identified was recorded (see Methods). Figure 4.2 provides the confusion matrix for the results of the procedure.

For both models, the recovery rate is approximately 85%. To account for errors in model assignment, in all of the following analyses, observations from a given participant are weighted by the difference in log-likelihood between the fit of the two models – that is, the stronger the evidence is in favor of one model, the greater the weight associated with the corresponding observation.

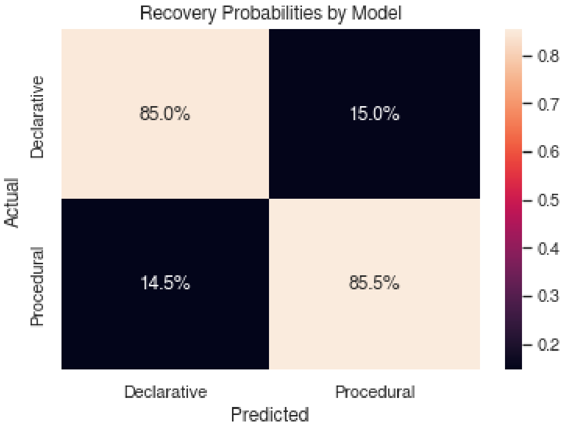

Fig 4.2. Confusion matrix showing the probability of correct model identification from 40,000 simulated runs. See Methods for details.

A second analysis was carried out to test whether each model's *parameters* (i.e., d and nu for the declarative model and alpha and tau for the procedural one) could also be correctly identified. This was done to make sure that any inferences about individual differences in the relative efficiency of the two systems could be grounded on reliable differences in parameters. Figure 4.3 illustrates the results of these simulations. All four parameters could be correctly identified from the simulated data, with correlations ranging from $r = 0.68$ (procedural temperature T, $p < 0.001$) to $r = 0.88$ (procedural learning rate alpha, $p < 0.001$).

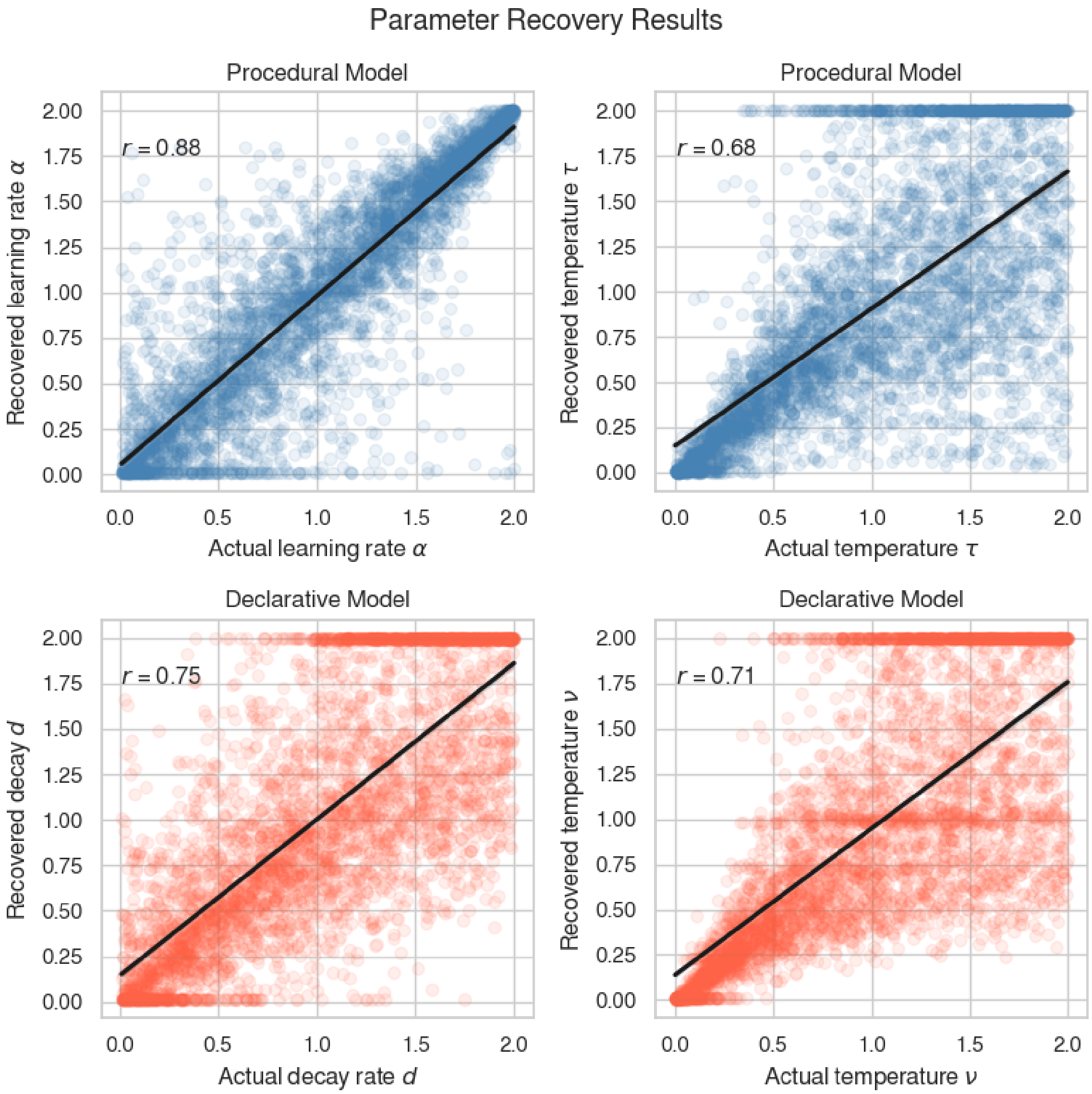

Fig 4.3. Results of the parameter recovery analysis for the individual parameters of the procedural (blue, top row) and the declarative (red, bottom row) models from 40,000 simulated runs. *See Methods for details.*

Decision-Making Process Identification. Having examined the reliability of the models, we then proceeded with categorizing participants based on which model fit them best. A total of

199 participants were fit by the two models. Of these, 35 (17%) were best fit by the declarative model and thus were included in the declarative group.

The remaining 164 individuals were best fit by the procedural model and included in the procedural group. Figure 4.4 illustrates the distributions of the log-likelihood differences (Procedural - Declarative) across participants. Individuals for which the difference was < 0 were categorized as procedural, and those whose difference was > 0 were categorized as declarative.

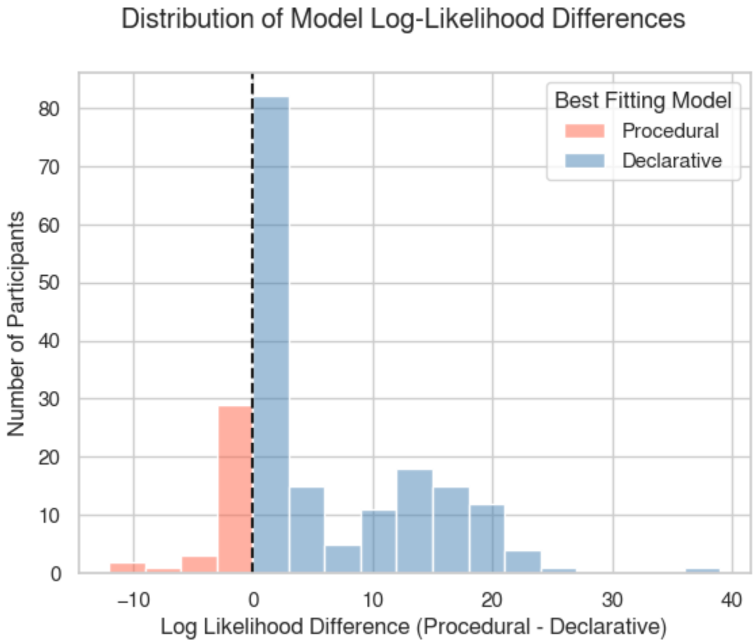

Fig 4.4. The distributions of the log-likelihood differences (Procedural - Declarative) across participants. *Red bars represent the count of participants who are best fit by the procedural model, whereas blue bars represent the count of participants who are best fit by the declarative model.*

<u>*Group Differences In Behavioral Responses During The Task.*</u> Our prediction is that, because of the different dynamics in the two models, individuals best fitted by the procedural model would exhibit greater sensitivity to trial feedback and a greater probability of switching

from their previous decision than individuals best fitted by the procedural model. To test this prediction, a dummy variable was created to code each trial after the very first as either a response switch (1) or not (0). The effect of the previous trial feedback on the probability of a response switch was then analyzed using a logistic mixed-effects model implemented in the lme4 package in R (Bates et al., 2015). In the model, Group (declarative vs. procedural) and Feedback (Win vs. Loss vs. Neutral) were treated as fixed effects, and individual subjects were treated as random effects, as shown in Figure 4.5.

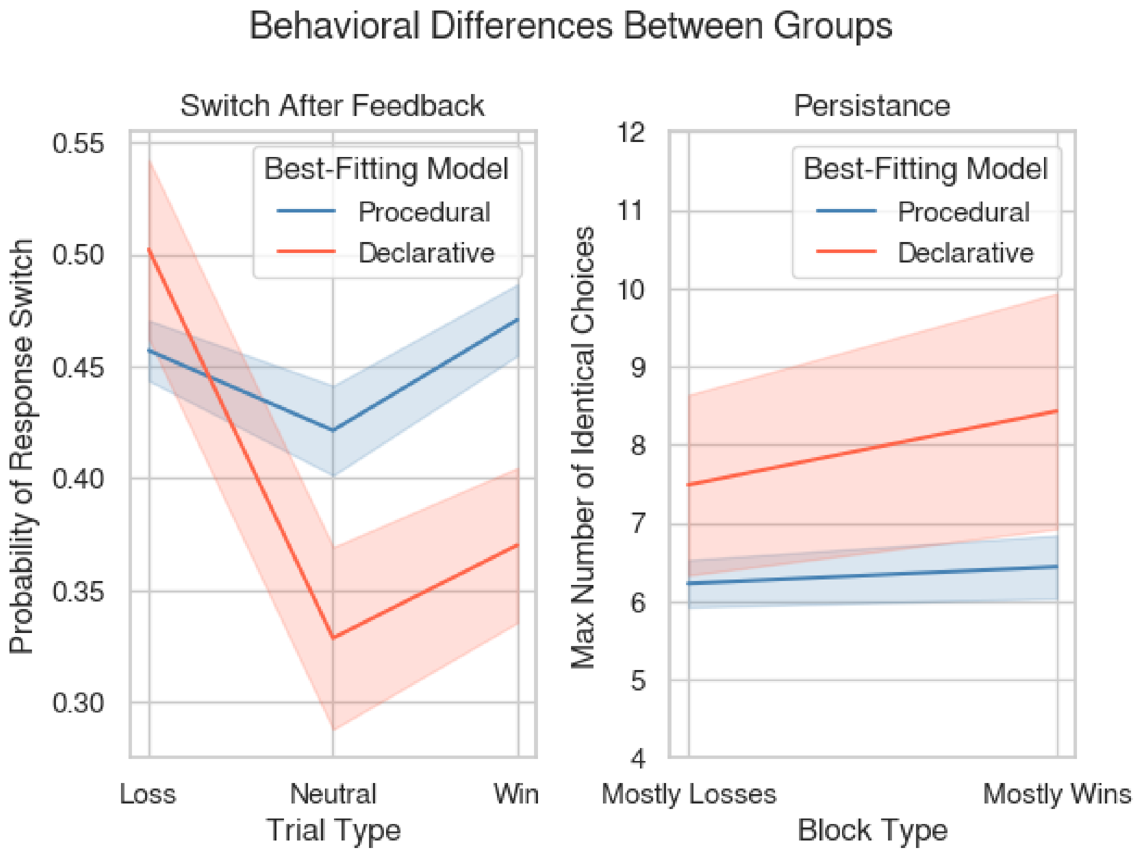

Fig 4.5. Main behavioral differences between participants best fit by procedural (blue) or declarative models (red). *Left: the procedural group shows a greater tendency to change responses after and neural and a "win" feedback. Right: the declarative group shows a greater tendency to repeat the same choice within a block.*

As expected, we found a significant main effect of model ($\beta = 0.87$, $t = 2.21$, $p < 0.02$), a significant main effect of trial when the feedback was a loss ($\beta = 1.13$, $t = 2.92$, $p = 0.003$), and a significant interaction between the two factors ($\beta = -0.95$, $t = -2.44$, $p = 0.01$). In summary, participants in the procedural group showed a significantly greater probability of shifting their behavior than participants in the declarative group, and, although both groups were more likely to change their responses after a “loss” than a “win”, this difference was magnified in the declarative group (Figure 5, left). This finding was consistent with our understanding of the memory retrieval process, with the declarative system becoming less sensitive over time to individual items of feedback in relation to the accumulation of historical decision/outcome pairs.

We also reasoned that the declarative group would exhibit longer streaks of identical choices than the procedural group. Furthermore, since both models are more likely to change their responses after a “loss” than a “win”, we expected that for both models the persistence would be larger in the “Mostly Win” blocks. To test these predictions, the maximum number of consecutive identical choices was computed for each block and individual, and the data were analyzed with a linear mixed model implemented in the *glmer* package in R. In the model, the persistence variable was modeled with a Poisson distribution, and both Block (“Mostly Lose” vs. “Mostly Win”) and Model (declarative vs. procedural) were included as fixed-effects. Because there is only one observation per participant (the maximum number of consecutive choices), the analysis used a simple linear model with no participant-level intercept and slopes. The results showed a significant main effect of both block (blocks: $\beta = 0.25$, $t = 5.56$, $p < 0.0001$) and group (β -0.70, $t = -19.64$, $p < 0.0001$), as well as their interaction ($\beta = -0.13$, $t = -2.78$, $p = 0.005$). As expected, the declarative group had higher persistence than the procedural group and this effect was magnified in the “Mostly Win” block (Figure 4.5, right).

Group Differences in Brain Activity During the Task. In addition to finding significant behavioral differences between two groups of participants, we also found relevant and consistent differences in their task-based fMRI data. In line with previous studies using the same dataset (Newell & Rosenbloom, 1982), our investigation focused on the difference in brain activity between “Mostly Win” and “Mostly Lose” blocks. This contrast is particularly important in this analysis, since participants are more likely to keep engaging their preferred decision-making system during “Mostly Win” blocks (as implied by Figure 4.5, right).

A *t*-test was performed to identify brain regions more active in one group over the other. The test was weighted using the relative log-likelihood of each participant’s being fitted by either of the two models, so that participants who had a stronger preference for one process were weighted more heavily (See Materials and Methods). The False Discovery Rate (Benjamini & Hochberg, 1995) procedure was used to correct for multiple comparisons at a level of $q < 0.01$. The analysis identified several brain regions that show significant BOLD signal differences between the procedural and declarative groups (Figure 4.6; Table 1). Specifically, participants in the procedural group (Figure 4.6, blue) showed greater task activation in the primary (inferior and superior occipital cortex and lingual gyrus) and secondary (left and right inferior temporal cortex) visual regions, and in the primary (M1) and secondary (supplementary motor area) motor regions, together with greater task activation in the bilateral insula. This pattern of activation is consistent with the acquisition of habitual stimulus-response associations.

On the other hand, participants in the declarative group (Figure 4.6, red) showed greater brain activity in the canonical nodes of the default mode network (the bilateral medial frontal cortex, the bilateral precuneus, and the bilateral inferior parietal lobules) as well as regions involved in the encoding (right hippocampus, left anterior temporal lobe), and the retrieval (left and right lateral ventral prefrontal cortex) of declarative memories.

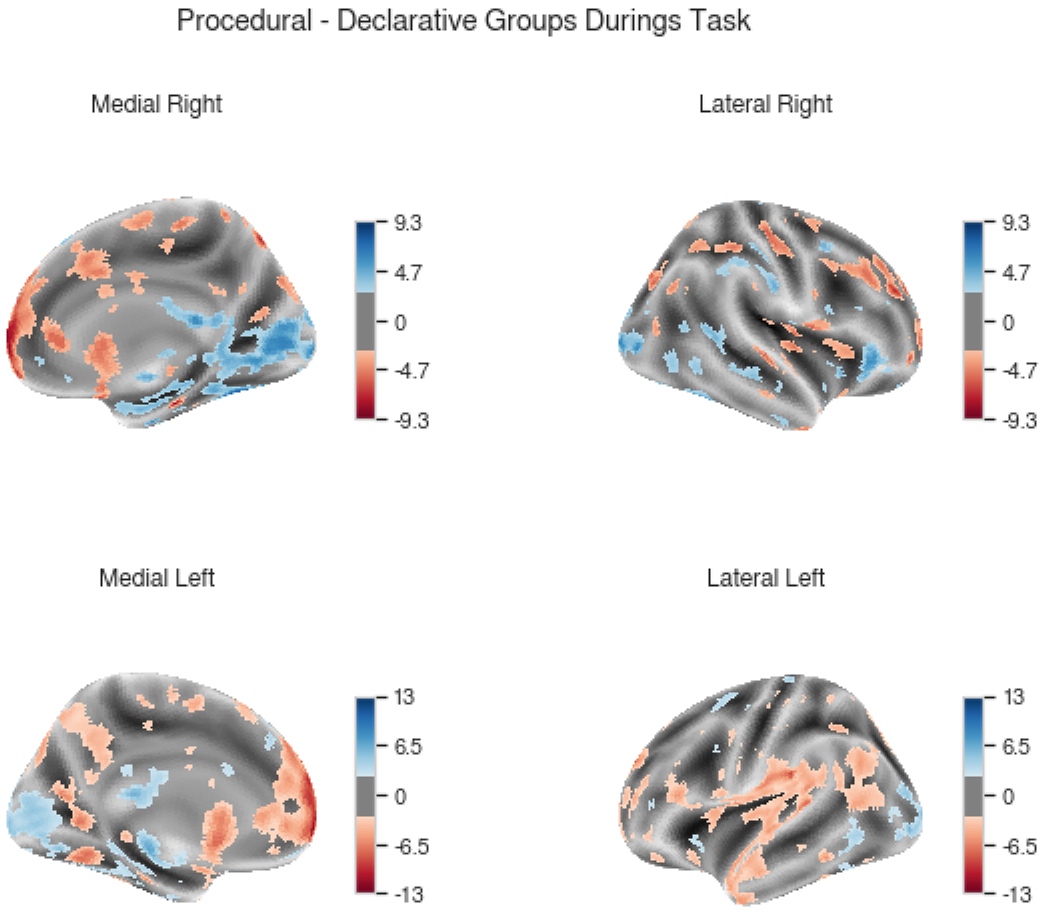

Fig 4.6. Differences between procedural and declarative participants during the incentive processing task. *In the overlays, the color indicates the direction of the difference (blue: Procedural > Declarative; Red: Declarative > Procedural) and color intensity indicates the magnitude of the associated T statistics. Regions are thresholded at a significance level of at $q < 0.01$, corresponding to a voxel-wise threshold of $T >= 2.70$.*

Table 4.1: *Results from the task fMRI comparison. For clarity, only clusters spanning > 50 voxels and locations spanning > 1% of the cluster and > 10 voxels within the cluster are shown. L = Left, R = Right; the word "gyrus" is omitted from the location names.*

Contrast	**MNI (Peak)**	**Automated Anatomical Labeling (AAL) Locations**	**Size (voxels)**	**Peak *T* value**
Procedural > Declarative	22, -82, -48	R Cerebellum (4.44%); L Cerebellum (5.22%); R Fusiform (3.15%); L Calcarine (2.99%); R Calcarine (2.94%); R Lingual (2.76%); L Middle Occipital (2.64%); R Middle Temporal (2.14%); Left Fusiform (2.12%); L Lingual (1.62%); R Inferior Frontal (1.35%); R Inferior Temporal (1.27%); R Middle Occipital (1.21%); R Hippocampus (1.13%); L Inferior Occipital (1.12%); L Inferior Temporal (1.01%)	30,662	14.09
	62, -26, 46	R Postcentral (34.73%); R Supramarginal (34.27%); R Inferior Parietal(17.60%); R Rolandic (11.89%)	858	7.48
	-14, -34, 58	L Paracentral (43.00%); L Postcentral (6.11%); L Precuneus (4.07%); L Precental (1.53%)	393	5.55
	-22, 4, 48	L Superior Frontal (36.72%); L Middle Frontal (35.94%); L Supplementary Motor Area (3.52%)	256	4.41
	-10, 36, 42	L Superior Frontal (90.04%); ; L Middle Frontal (2.90%)	241	5.23
	36, 14, 60	R Middle Frontal (96.94%);	196	6.75
	12, 24, 20	R Anterior Cingulate (57.05%);	149	5.49
	-6, 18, 26	L Anterior Cingulate (40.74%); L Median Cingulate (10.37%	135	4.9
	-40, 16, -4	L Insula (92.06%)	126	5.1
	10, -22, 66	R Paracentral (43.93%); R Supplementary Motor Area (29.91%); R Precental (26.17%)	107	4.73
	-24, 68, 16	L Middle Frontal (24.14%); L Superior Frontal (18.39%)	87	4.61

	-40, -8, 52	L Precental (91.89%)	74	4.15
	28, 62, 24	R Middle Frontal (56.16%); R Superior Frontal (39.73%)	73	5.47
	-32, 0, 2	L Putamen (56.72%)	67	4.33
Declarative > Procedural	-4, 66, -2	L Superior Frontal (7.26%); L Superior Temporal (3.29%); L Middle Temporal (3.25%); R Superior Frontal (5.56%); L Precuneus (2.12%); L Middle Frontal (2.12%); L Postcentral (2.00%); L Inferior Temporal gyrus (2.74%); L Inferior Frontal part (1.58%); L Angular (1.54%); L Rolandic (1.47%); L Anterior Cingulate (1.43%); L Supramarginal (1.35%); L Caudate (1.35%); R Median Cingulate (1.27%); R Superior Temporal (1.24%); L Precental (1.17%); L Medial Frontal Cortex (1.13%); R Insula (1.10%); L Insula (1.06%); R Anterior Cingulate (1.05%); R Supplementary Motor Area (1.04%); L Temporal Pole (1.03%); L Supplementary Motor Area (1.01%); R Precuneus (0.99%)	38,651	-15.92
	22, -44, -54	R Cerebellum (86.87%)	545	-15.02
	22, -24, -26	R Parahippocampal (26.42%); R Cerebellum (33.97%); R Fusiform (18.24%)	159	-12.17
	64, -4, 32	R Inferior Frontal (38.31%); R Postcentral (31.17%); R Precental (30.52%);	154	-9.32
	40, 0, -44	R Inferior Temporal (89.47%); R Temporal Pole (6.58%)	152	-7.13
	2, -88, -16	L Calcarine (25.49%); L Lingual (17.65%); L Crus (7.84%);	102	-8.99
	6, -20, 76	R Paracentral Lobule (59.60%); R Supplementary Motor Area (12.12%);	99	-7.52
	52, -38, 56	R Postcentral (89.53%)	86	-8.25
	0, -4, 32	R Median Cingulate (52.56%); L Median Cingulate (39.74%)	78	-4.21
	-42, 20, 50	L Middle Frontal (98.68%)	76	-6

	14, -26, -8	R Lingual (15.71%); R Hippocampus (8.57%);	70	-4.72
	40, 24, -26	R Temporal Pole (98.33%);	60	-5.2

Predicting Individual Differences in Decision-Making Processes Through Functional Connectivity. Functional connectivity data consisted of pairwise partial correlation matrices between each pair of the 264 regions in the Power parcellation scheme (Power et al., 2011). To identify which functional connectivity features predict whether a participant would belong to the declarative or procedural group, we used a logistic Least Absolute Shrinkage and Selection Operator (LASSO: Tibshirani, 1996) regression model, which reduces a large number of potential predictors while retaining the most predictive features. The LASSO model was implemented using the *glmnet* package in R (Friedman et al., 2010).To ensure generalization, we chose the optimal LASSO hyperparameter λ_{nested} (0.0059) using a nested cross-validation method (See Supplementary Materials). Having the optimal λ_{nested}, we refit the LASSO model using the Leave-One-Out (LOO) cross-validation approach and obtained the correlation coefficient betas (β) for subsequent brain connectivity analysis.

The prediction performance was evaluated by regular LOO cross-validation and nested cross-validation approaches. The training and testing accuracy with regular cross-validation and nested cross-validation approaches are shown in Table 4.2, and The training and testing area under the (AUC-ROC) curves are shown in Figure 4.7. These results indicate that the LASSO model was successful at predicting an individual's preferred decision-making process (declarative or procedural) from resting-state brain functional connectivity.

Table 4.2. *Summary of the LASSO Model Performance*

	Best λ	Testing ROC-AUC	Number of Non-Zero β
Regular CV	0.0027	1.000	92
Nested CV	0.0059	0.911	91

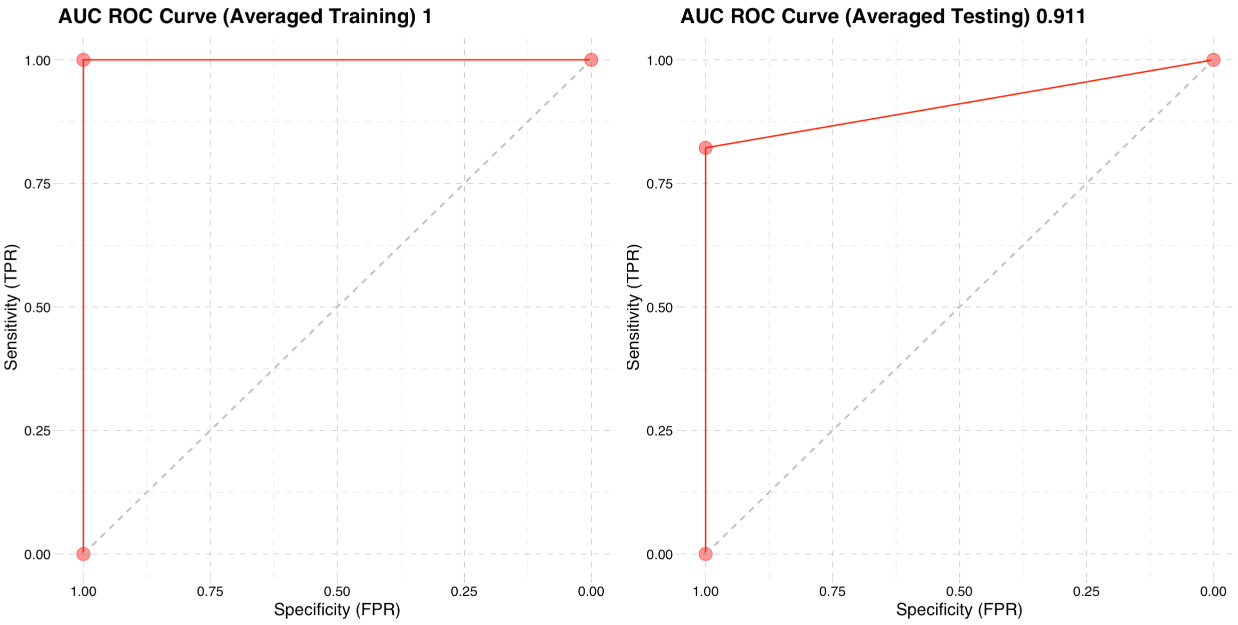

Fig 4.7. The Training and Testing Area under the ROC Curve*. (Left) Averaged training ROC with nested cross validation. (Right) Averaged testing ROC with nested cross validation.*

Finally, we examined which functional connectivity features were predictive of individual preferences in decision making. After LASSO regularization, 92 functional connections (approximately 0.11% of the total) had non-zero β parameters, suggesting a very sparse neurofunctional connectivity. Figure 4.8 shows the non-zero β weights of connection between networks and within networks as extracted from the fit LASSO model. Since the ultimate effect of a β parameter on the predicted group assignment depends on the polarity of the underlying functional connectivity, a positive β value has different implications if applied to a positive or negative partial correlation between two regions. Consider, for example, a connectivity value associated with a positive β value: if the underlying connectivity is positive ($r > 0$), then the degree of functional connectivity can be taken as a vote in favor of the declarative system. If, on the other hand, the underlying connectivity value is negative ($r < 0$), then it should be counted as a vote for the procedural system.

To make the interpretation of the values unambiguous, we examined the corrected weight matrix by multiplying the β matrix with the sign of the functional connectivity matrix, obtaining a corrected weight correlation matrix W. The weight matrix W can now be interpreted unambiguously, since $W > 0$ predicts reliance on a declarative process and $W < 0$ predicts reliance on a procedural process. Figure 4.8(A) shows these different functional connections.

Connections predictive of declarative processes involve clearly different networks than connections predictive of procedural processes, as shown in Figures 4.8. As expected, reliance on declarative processes in decision-making was predicted by greater connectivity in the networks of regions associated with task control (task control networks and subcortical networks) as well as episodic memory (memory retrieval networks). Reliance on procedural learning processes, however, was predicted by greater connectivity in sensorimotor and salience regions. From Figures 4.7 (A), we could find that the predictive functional connectivity also largely overlaps with the task-fMRI results, indicating that greater functional connectivity between two regions at rest predicts a greater likelihood that these two regions would be used to support the decision-making processes during the task.

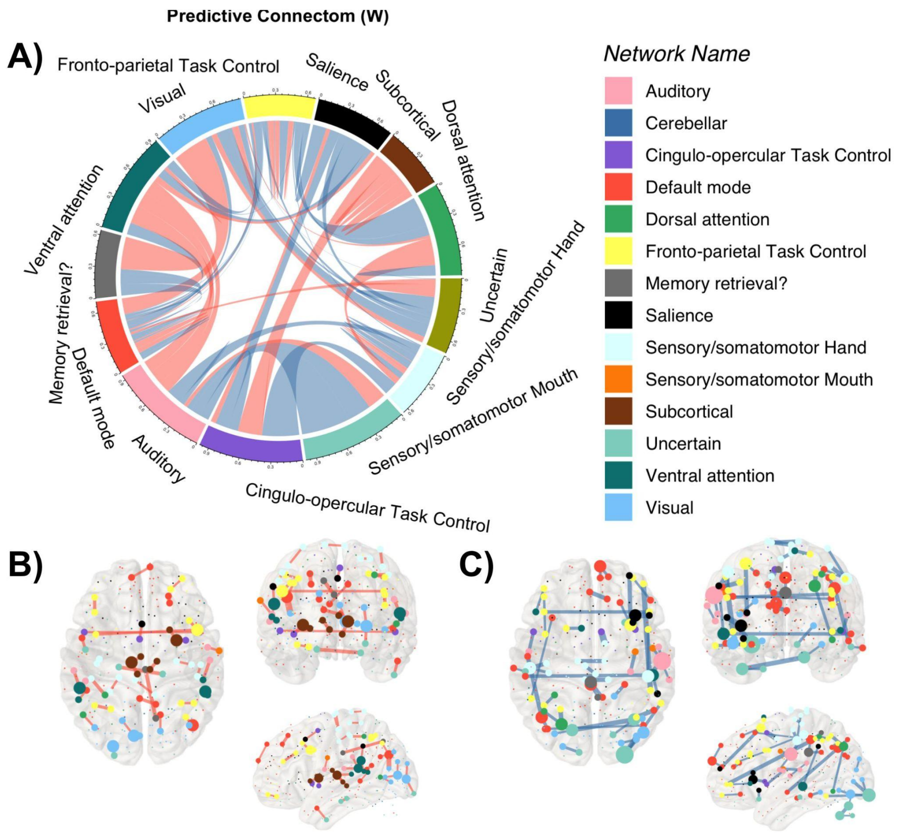

Fig 4.8. The distribution of predictive connections in declarative and procedural functional networks. *(A) Predictive Connectom across the Power parcellation networks. Red connections denote positive weight values (W > 0), predicting the procedural Group, and Blue connections denote negative weights values (W < 0), predicting the procedural Group. (B) The anatomical location of the function connection is predictive of declarative group assignment; nodes represent regions, node colors indicate the network they belong to; node size indicates the degree of connectivity (the number of adjacent edges), and line width indicates the strength of connectivity between two nodes. (C) Anatomical location of the functional connections predictive of procedural group assignment.*

4.6 Discussion

This study shows that individuals rely on different mechanisms when deciding from experience, and that this preference is adaptive and reflects individual differences in the functional connectivity between each process' corresponding circuitry. Specifically, individuals exhibiting stronger connectivity between and within frontoparietal and memory retrieval regions tend to use declarative strategies that are more reliant on episodic encoding and retrieval, while individuals with stronger connectivity in cingulate, sensory, and basal ganglia regions tend to rely on habitual actions and reinforcement learning. An individual's preference can, in fact, successfully be predicted from their underlying functional connectivity.

Although our results shed new light on the neural bases of experiential decision-making, a number of limitations must be acknowledged. First, participants were assigned to the declarative or procedural group based on the log-likelihood of a corresponding cognitive model. Because no other ground-truth labels were available, these classifications should be interpreted with caution. Another limitation is that the incentive processing task differs from most decision-making paradigms, as there is no winning rule for participants to learn from the feedback. Thus, although this task was particularly well suited for the current study, more work is needed to determine whether these findings would translate to more realistic situations.

In addition, our study assumes that greater functional connectivity reflects greater efficiency. While much experimental evidence points to this, the mechanisms by which connectivity translates to computational efficiency are not clear. One proposed solution is that connectivity reflects communication efficiency across regions. If so, connectivity should correspond to decision-making noise in our models. In fact, a follow-up analysis of individual model parameters shows that individual preferences for one process over the other largely follow differences in the estimated noise parameter. That is, individuals whose estimated procedural noise was larger than the estimated declarative noise were significantly more likely to be fit by the declarative model, and vice-versa (Welch Two Sample *T*-test: $t(50.04) = 3.37$, $p = 0.001$, see Figure 4.9).

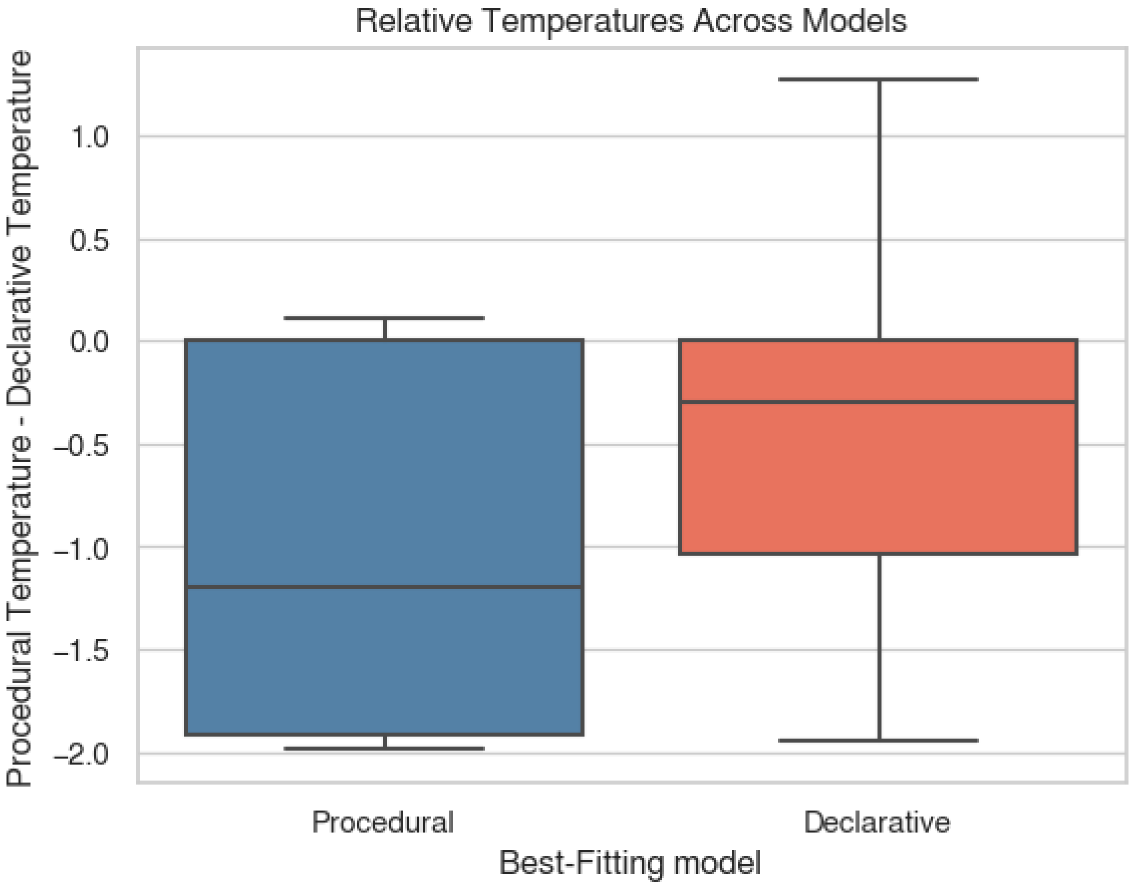

Fig 4.9. Estimated procedural and declarative temperature parameters between the two groups.

Although we have found evidence that individuals prefer to implement decision-making processes using the circuits that show the greatest connectivity, our analysis does not directly support a causal direction. To conclusively show a direct causal effect, it would be necessary to experimentally alter functional connectivity within the declarative and procedural circuits and measure subsequent changes in behavior in the same task. Such an experimental intervention is difficult to carry out in humans, although it is in principle conceivable to use pharmacological interventions that directly target one circuit or intracranial direct stimulation of these circuits in patients with implanted electrodes or ECoG grids.

Finally, our study still does not directly address the source of differences in functional connectivity. Two hypotheses are possible. In principle, these differences could reflect underlying genetic or structural differences in brain function. Alternatively, it is possible that differences in functional connectivity simply reflect one's past history of relying on one particular process (and its underlying circuit) over the other. In this case, individuals might have a preference (let's say, for the declarative system), and repeated use leads to corresponding increases in functional connectivity (let's say, in the memory retrieval network). The two explanations might be intertwined, i.e., small initial differences in the efficiency of these circuits might lead to greater use of one process over the other, and practice over time might result in additional gains in efficiency.

Further investigation is needed to fully untangle these questions, but these results nevertheless represent a promising step forward in directly connecting the mechanical underpinnings of brain circuitry with observed human behavior in decision-making.

Chapter 5. STUDY 3: THE COST OF MODEL-BASED DECISION MAKING IS ROOTED IN MEMORY [2]

5.1 ABSTRACT

Understanding the fundamental cognitive process of decision-making is crucial for developing appropriate cognitive models. Two main planning-based approaches have been used to investigate learning in complex decision-making tasks: one using model-based reinforcement learning (MB-RL), an extension of reinforcement learning that includes high-level planning, and the other using instance-based learning (IBL), based on episodic memories of previous interactions. In this paper, we attempt to reconcile the two approaches by using ACT-R to implement a cognitively plausible substrate for the planning component of MB-RL. We review the model-based (MB) and model-free (MF) learning approaches in reinforcement learning and discuss their roles in decision-making strategies. Within the ACT-R framework, we propose a promising model that incorporates memory retrieval in MB, offering a cognitively plausible approach to the planning component of MB-RL. Our combined model successfully replicates well-known findings in the literature, including developmental reliance on memory and response time variations between common and rare options. Finally, our model naturally accounts for the balance of memory and RL depending on the relative cost of each. We argue for the superiority of our cognitive model and address the significance of this study for understanding the brain and computational processes underpinning decision-making strategies, as well as for applications in artificial intelligence and decision-making modeling.

---

[2] A reduced version of this chapter has been accepted for the 2023 International Conference on Cognitive Modeling.

5.2 BACKGROUND

Decision making is a fundamental ability of human cognition. Extensive research has been conducted on the mechanisms of experiential decision making in humans and animals. The predominant view is that, under simple circumstances, decisions are well characterized by model-free reinforcement learning (MR-RL). In MF-RL, the decision maker is an agent, and the available options are actions that the agent can apply to an environment. The agent typically uses temporal difference (TD) methods, such as Q-learning, to improve the estimates of future rewards associated with each action.

The MF paradigm has been extremely successful at explaining both behavioral and neural data in animal and human experiments (Niv, 2009). Most decisions, however, are not made within the simplified boundaries of laboratory experiments. This is particularly true in the case of humans, who interact with complex, non-stationary environments.

To deal with more complex situations, researchers have borrowed the concept of model-based reinforcement learning (MB-RL), an extension of reinforcement learning that includes additional memory structures to explicitly store changes in the environment following an action from the agent. The MB approach involves the construction and use of an internal representation of the environment, which allows for flexible and goal-directed decision-making. MBRL is a heterogeneous collection of methods, some of which include explicit replay of previous experiences (Sutton, 1991) while others are purely planning algorithms (Glascher et al, 2008).

Cognitive Substrates of MB. Despite much research, it is still unclear what cognitive processes underlie MB learning; some authors refer to it as "planning," while others link it to memory. Additionally, there is evidence that the MB and MF strategies are frequently combined, but there are no established standards for figuring out the best combination of these two approaches, especially in the context of cognitive literature.

Doll et al. (2015) provide critical insights into the interplay between MB and MF reinforcement learning approaches in decision-making strategies. They pointed out that the brain's multiple memory systems, specifically the declarative and procedural memory systems, serve as crucial substrates for distinct decision systems. Declarative memory, which involves conscious recollection of facts and events, is associated with MB learning as it enables the construction of mental models of the environment and the planning of actions based on simulating potential action outcomes. In contrast, procedural memory, which entails learning habits and skills, is linked to MF learning, where decisions are guided by learned associations between actions and outcomes. Doll et al.'s findings further provide insights into how these two memory systems are dependent on each other in learning decision-making strategies, highlighting the complex and adaptive nature of human cognition.

Instance-Based Learning. When dealing with decisions in complex tasks, a radical alternative to MB is the hypothesis that humans rely on memories of previous interactions. Perhaps the most promising part of this framework is the Instance-Based Learning (IBL) theory. Gonzalez, Lerch, and Lebiere (2003) pioneered this line of inquiry with their IBL framework, which integrates elements of both MB and MF learning approaches. They present a cognitive model within ACT-R to explain how people make decisions in dynamic environments. They proposed that humans memorize specific instances of their interactions, such as the action taken and the associated outcomes, and use these memories to inform future decisions. When confronted with a new decision, individuals retrieve the most typical instance(s) from memory and use this instance's actions and outcomes to guide their current choice. This process is influenced by the perceived utility of past actions and the similarity between the current situation and stored instances. As individuals accumulate more experiences from the environment, their decision-making processes become more refined and better aligned with the changing environment.

IBL mechanisms share similarities with the MF approach in RL, as both rely on learned associations between actions and outcomes without explicitly constructing an internal model of the environment. In MF learning, agents learn through trial and error, gradually adjusting their behavior based on the rewards received from previous actions. Similarly, IBL emphasizes the role of past experiences in shaping decisions, with individuals using stored instances to inform their choices.

On the other hand, IBL can also be connected to MB-RL, particularly in the context of the ACT-R framework. In MB learning, agents plan actions by simulating the consequences of different choices. While IBL does not explicitly build a complete model of the environment, it relies on memory retrieval and the evaluation of previously encountered situations to inform decision-making. This aspect of IBL aligns with the cognitive processes in the MB approach, such as working memory and planning. Unlike MBRL, IBL makes specific predictions about which cognitive and neural resources will be used. Unlike MBRL, IBL includes no explicit planning and no explicit encoding of state dependencies. In this sense, it approximates MF more than MB.

Present Study. In this paper, we aim to integrate the classical approach to MB with the insights gathered from the IBL use of long-term memory to guide decision-making. We propose an integrated cognitive model that relies on declarative long-term memory to implement MB learning, and uses ACT-R's declarative model to give cognitively plausible implementation of these operations. We argue for the superiority of our cognitive model over the traditional RL model and discuss the implications of this research for understanding the neural and computational mechanisms related to cost-benefit evaluation underlying decision-making strategies, as well as for applications in areas such as artificial intelligence and decision-making modeling.

5.3 METHODS

Dataset. Nussenbaum et al. (2020) conducted an online experiment using the Markov two-stage task paradigm to replicate the main findings from Potter et al. (2017) and Decker et al. (2016) that MB learning increased as age increased. In this task, participants aimed to gather a "space treasure" by exploring various planets inhabited by different alien treasure-miners. They made first-stage choices between two spaceships, each with different probabilities of reaching specific planets, and subsequently, they made second-stage choices on each planet, asking one of two aliens for treasure based on slowly drifting reward probabilities.

The de-identified behavioral data is obtained from the Open Science Framework by Nussenbaum et al., 2021: https://osf.io/we89v/. A total of 151 participants (fifty children; fifty adolescents, and fifty-one adults) were included in this study. The computational model in ACT-R was developed using a similar paradigm, but with abstract stimuli (A1 and A2 as two options at stage 1; B1 and B2 as two options at stage 2, representing state B, C1 and C2 as two options at stage 2, representing state C). All participants included in the study were healthy individuals without any neurodevelopmental or neuropsychiatric disorders. The experimental protocol, subject recruitment procedures, and consent to share de-identified information were approved by the Institutional Review Board at Washington University.

The Two-Stage Task. The Markov two-stage task paradigm (Figure 5.1), has been used to study a wide range of cognitive processes, including learning, decision-making, and memory. In this task, participants are presented with a series of trial screens, referred to as "states" and indicated as A, B, and C. Each state contains two options, indicated as A1 and A2; B1 and B2; C1 and C2. Participants are asked to select one of the two options using the keyboard left or right. They always start in state A and, depending on the option they choose, will transition to state B or C. Their choices in the second state determine their choice of receiving a reward.

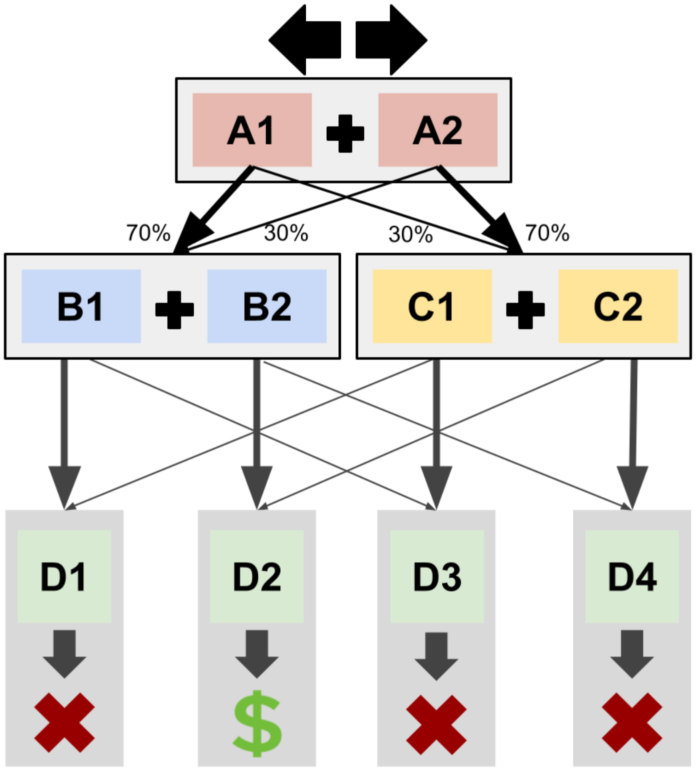

Fig 5.1. Two-Stage Task Paradigm. *Black arrows indicate the state transition probability, the thick line is 70% and the thin line is 30%. One of the final states is associated with a reward, and the probability of receiving a reward changes slowly across the experiment following a random walk with a mean of 0 and a standard deviation of 0.025.*

The probability of a state transition from selecting one option in the first state to a specific second-stage state is predetermined, according to 70% of the time, and this state transition probability remained consistent throughout the entire experiment. For example, selecting A1 will have a 70% chance of resulting in the B state, referred to as the common state frequency, and a 30% chance of resulting in the C state, referred to as the rare state frequency. Similarly, selecting A2 will have a 70% chance of resulting in the C state (common) and a 30% chance of resulting in the B state (rare). To promote continuous learning, the probability of receiving a reward for selecting the sequence of options slowly changes on each trial following Gaussian random walks (*M* = 0, *SD* = 0.025).

The task consists of two phases: a learning phase and a choice phase. Each stage is divided into three blocks. In the first two blocks of the learning phase (*N* = 20), participants are free to explore the transition probability between states by randomly selecting one of the two options, but no reward will be given in the end. After a short break, participants will be able to collect rewards at the end state with a slowly changing probability. This block is designed to allow participants to potentially learn the relative value of a sequence of choices by experiencing the outcomes. The first two blocks of the choice phase are identical to the learning phase, but the last reward block consists of *N* = 201 trials. Participants must rely on the information they learned during the learning phase to make their selections.

MB and MF Patterns in the Two Stage Task. The two-stage task was developed to separate the contributions of MB and MF learning to decision-making. To understand how this is possible, one must consider two factors. The first is that, after feedback is delivered, the values of the actions that led to it will be updated accordingly. This means that the probability of repeating the same initial action in the same trial, indicated as the *Stay Probability*, will change.

The second consideration is that MF and MB update the values of the actions in different, and sometimes opposite ways. MF learning is blind to the circumstances that lead to the reward and will simply increase the value of the preceding actions. The MB learning system, by contrast, has access to information about the state transition probabilities and can update the values of actions based on them.

Specifically, if a reward was delivered after a first-stage action that led to a rare state transition such as when the agent selects “left” (A1) at stage 1 and reaches state C at stage 2 (which occurs with a 30% chance), the MB learning would prefer to increase the value of the opposite action (“right” action at stage 1), since that has a greater likelihood of leading to the rewarded state. This, in turn, would lead to a decrease, rather than an increase, in the stay probability. In other words, while the stay probability is affected by both reward and transition frequency, the stay probability shows an interaction of reward and transition frequency in MB, with the reward having opposite effects on actions that lead to common or rare transitions. Figure 5.2 illustrates the prototypical behavior of MF-RL, MB-RL and Hybrid RL models in this task.

Multiple empirical studies revealed that human participants demonstrated a mixture of MF and MB that combined elements of both MF and MB strategies (Gläscher et al., 2008; Daw et al., 2011; Otto et al., 2013). By analyzing the probability of staying with the same first-stage option as a function of reward and transition frequency, researchers can infer the extent to which participants rely on MF, MB, or hybrid learning strategies in the two-stage task.

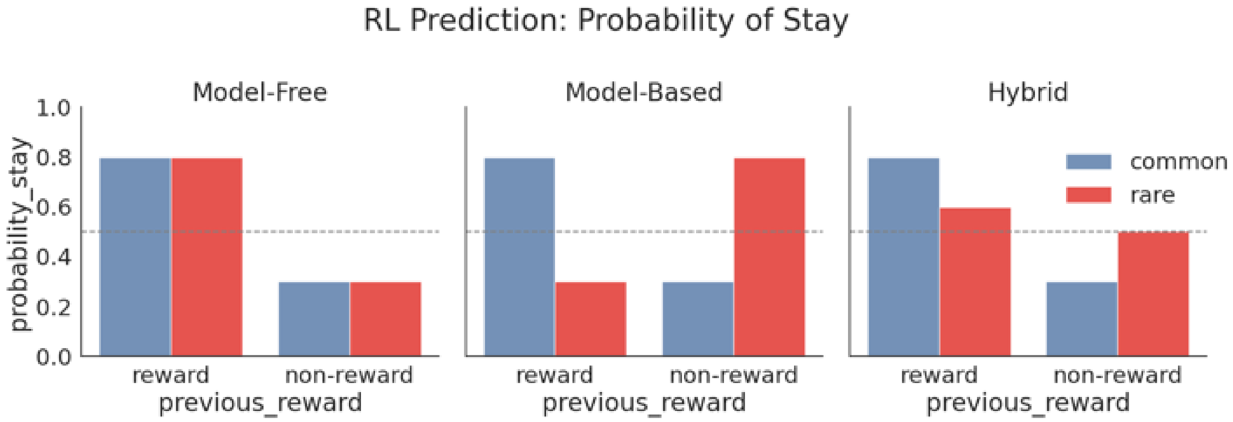

Fig 5.2. The canonical RL prediction of the probability of staying as a function of reward and transition frequency. *(A) RL Model-Free (B) RL Model-Based (C) RL Hybrid*

5.4 COMPUTATIONAL MODELS

In the next section, we will illustrate two different model implementations capable of performing this task. The first is a "pure" RL model that combines MF and MB components and includes no specific substrate for the MB component. This is the de-facto implementation that is commonly used across multiple research papers that have used this task (Daw et al., 2011; Doll et al., 2015; Kool et al., 2016; Gershman & Daw, 2017; Weissengruber et al., 2019). The second will be a hybrid model that integrates ACT-R's declarative model within the MB component. The latter implementation aims to develop a more cognitively plausible and robust framework that can better simulate human cognitive processes in a dynamic and noisy environment.

5.4.1 *Pure RL Model*

MF Component. The central idea of MF learning is to gradually update the value of actions by calculating the difference between expected and actual rewards gained from the environment. The expected value of state-action is calculated based on the SARSA temporal difference algorithm using Eq 5.1 (Daw et al., 2011; see Chapter 2 Eq 2.2), where α denotes the learning rate, r denotes the immediate reward received after taking action a in state s; the discount factor, denoted by gamma (γ), determines the importance of future rewards compared to immediate rewards; and (s, a) refers to state-action at current state, (s', a') refers to state-action at the next state.

$$Q(s', a') = Q(s, a) + \alpha\delta_{i,t}$$

where

$$\delta_{i,t} = r + \gamma Q(s'_t, a'_t) - Q(s_t, a_t)] \qquad (5.1)$$

<u>MB Component</u>. Unlike MF, which learns by trial-and-error, MB Learning is based on the idea that agents could create a mental representation of the environment and plan accordingly. Here, "planning" means simulating various future trajectories. This is achieved by predicting the consequences of possible actions and then using these predictions to select optimal actions.

The Q-value in MBRL is calculated and updated based on the Bellman equation (Eq 5.2). Similarly, the parameter r denotes the immediate reward received after taking action in state *s*; the discount factor, denoted by gamma (γ), determines the importance of future rewards compared to immediate rewards. $\Sigma s'$ is a sum over all possible next states, and max[Q] represents the maximum Q-value over all possible actions in the next state s'. P(s'|s,a) is the transition probability that agents have knowledge about the dynamics of the environment. In this formulation, the transition probabilities are fixed (0.3/0.7) and given as ground truth.

Using this equation, the Q values are gradually updated, converging to the optimal Q which is the expected reward for all states and all action pairs under the best policy (see Chapter 2 Eq 2.2).

$$Q(s,\ a) \ = \ r \ + \gamma \sum^{s'} P(s' \mid s,\ a) \cdot \ \max\ [Q(s',\ a)] \tag{5.2}$$

<u>*Combination of MF and MB Components*</u>. Empirical findings suggest that human subjects tend to adopt hybrid approaches rather than pure MF or MB learning (Daw et al., 2011; Decker et al., 2016; Otto et al., 2013). Empirical findings suggest that the interaction between these learning strategies is crucial in determining decision-making behavior, with the balance between the two being influenced by the task's state and transition structures.

One common approach to combining MF and MB learning approaches is to use the weighted sum of the MB and MF value estimates to make decisions. A weight parameter w determines the relative importance of the MB and MF estimates. In these models, the agent starts by using a MB strategy to plan its actions, but as it gathers more information about the environment, it shifts towards a MF strategy. This allows the agent to quickly adapt to changes in the environment while still being able to plan its actions efficiently.

In pure MF learning, the weight parameter is set to 0, while in pure MB learning, it is set to 1. In a hybrid model, the weight parameter can be set to a value between 0 and 1, depending on the task and the available information. The weight parameter can be learned through a process called “weight adjustment,” where the agent adjusts the weight based on the performance of the current strategy. This allows the agent to adapt to the changes in the environment and find the optimal balance between exploration and exploitation. However, the neurobiological meaning of this parameter is less clear, especially from the perspective of decision-making and its underlying cognitive processes.

5.4.2 *Hybrid ACT-R RL Model.*

MF Component. The MF component of the hybrid model is identical to the MF component of the pure RL model. In turn, this component is also broadly consistent with ACT-R’s procedural knowledge module, which also uses RL to learn stimulus-response associations in the form of procedural rules. Thus, we employed the SARSA MF framework mentioned above as a substitute for ACT-R's procedural module.

MB Component. RL-MB learning is creating a model of the environment, which allows the agent to plan its actions by simulating the consequences of different choices. Critically, it depends on the knowledge about transition probability, that is, how likely it is to move from the initial state to the next, given a particular action. In the pure RL model, these probabilities are directly provided to the model. However, this assumption may not fully capture the nature of learning and cognitive processes in the task. This knowledge is not simply given, but actively updated and accumulated by agents from the environment through their interactions with the external world.

Thus, in the hybrid model, the MB component encodes its knowledge of the environment as episodic long-term memories. Much like in the IBL approach, the model retrieves and inspects these traces as part of its planning process. The model’s long-term memory was developed using PyACTUp (Yang, Y., Morrison, D., Stocco, A., Orr, M., & Lebiere, C. 2020), a Python implementation of ACT-R’s declarative system. We argue that our ACT-R model performs as well as RL models in simulating canonical MB behavioral patterns, and even more importantly, it provides a plausible cognitive framework to understand how our brain represents the planning process.

Specifically, as the agent observes the two options, a two-step process of planning begins. The agent tends to retrieve the most likely (most frequent and recent) subsequent states given two possible actions (‘f’ and ‘k’) from the declarative memories. Based on the retrieved next state s’, the MB Q-value of two possible state 2 actions is calculated based on the Eq 3.

Additionally, unlike RL, our declarative model would estimate the transition frequency based on prior memories of trials. It samples the memories about state transitions and calculates the probability of state1-state2 given an action. That is, among the retrieved chunk samples, the estimated transition probability, P(s'|s, a), is calculated by the number of first state (A) to second state (e.g. B or C) divided by the total number of sampled memories given a particular action a. Here, the number of sampling times is fixed at an arbitrary number of 20. Although we did not investigate further how the number of sampling counts affects decision-making, it is reasonable to hypothesize that a larger n suggests a more accurate assessment of transition frequency and may indicate a greater effort in cognitive control planning or an individual’s greater working memory (WM) capacity. After estimating P(s'|s, a), the Q-values are computed as:

$$Q(s,\ a) = P(s'_1|s,\ a)\cdot\ \max\,[Q(s'_1,\ a)]\ +\ P(s'_2|s,\ a)\cdot\ \max\,[Q(s'_2,\ a)] \qquad (5.3)$$

This method has a significant advantage in approximating realistic transition frequencies for decision-making processes because it uses declarative memory to make educated estimates about the frequency of state-action pairs. In contrast to conventional MBRL algorithms, which typically assume a fixed 0.3/0.7 transition frequency as the basis for Q-value calculations, our ACT-R model incorporating declarative memory provides a more realistic representation of human cognitive processes. Because it is built upon a reliable model of memory, the estimates made by the agent reflect some of the distortions and fallacies of humans.

At the end of each trial, the agent forms a new episodic memory of the interaction, containing the states and actions taken. A new chunk, consisting of 5 slots: ‘stage’, ‘current_state’, ‘next_state’, ‘response’, and ‘reward’. is created and merged into existing Long-Term memory.

This two-step planning at stage 1 allows the agent to retrieve the most rewarding action sequence based on prior learning experience, taking reward and state transition frequency into account. Most importantly, the IBL in ACT-R could estimate the transition frequency from state1 to state2 using activation calculations, as shown in Eq 5.4 (See Chapter 2 Eq 2.4, Eq 2.5).

$$A_i = B_i + \varepsilon \text{, where } B_i = ln(\sum_{j=1}^{n} t_j^{-d}) \quad (5.4)$$

The calculation is based on the idea of time decay, where more recent uses contribute more to activation than less recent ones. The decay rate parameter determines how quickly the contribution of past memories diminishes over time. A higher decay rate leads to more rapid forgetting, emphasizing the role of recent interactions, while a lower decay rate allows the influence of older interactions to persist longer.

In addition to base-level activation, noise plays a pivotal role in the ACT-R’s declarative memory framework. Noise, denoted by ε in Equation 5.4, adds a level of randomness to the activation formula, which lets brain processes be different and hard to predict. This lets the model better represent the wide range of actions seen in actual data. When noise is added, chunks with less activity are more likely to be chosen. This encourages exploration and could lead to the discovery of better tactics.

The state2 planning is simpler than state1, since it consists of only one-step planning, with more restricted information. After attending to and encoding the state2 stimuli into the working memory, the agent sends a retrieval request asking for a state2-stimulus chunk that matches the state1 slot value. For example, if an agent chose to leave at state 1 and end up at state B. The state2 plan is searching any chunk that contains the current state and is equal to B. If the retrieved memory has a reward greater than 0, then choose this action as the state 2 response, otherwise, choose the alternative action.

Similarly, the most frequent and recent state2-stimulus will be retrieved based on the activation calculation, allowing the agent to choose the most available memories (s', a') by observing action-outcome associations from prior trials.

5.4.3 *Optimization with Maximum Log-Likelihood.*

Both of the models were fitted to each individual using maximum log-likelihood approach, which is a standard method used to estimate individual subjects' data in cognitive modeling research (Yang, Karmol & Stocco 2021). This method involves calculating the log-likelihood function, which measures the goodness of fit between the observed data and the hypothesized probability distribution. Using the softmax choice equation, as shown in Eq 5, we calculated the probability of selecting a specific response (either left or right) given the model parameters to estimate individual subjects' data using the log-likelihood approach.

$$P\ (a \mid s) = \frac{e^{\beta \cdot Q(s,\, a)}}{\sum_{a'} e^{\beta \cdot Q(s',\, a')}} \qquad (5.5)$$

We adopted a simpler parameter estimation pipeline, as in Decker et al. 2016; Potter et al. 2017; Nussenbaum et al., 2021. Three Q-learning parameters (α, β and λ, w), and two memory relevant parameters (temperature and decay) are fit to each individual subject. α is the learning rate in Q-Learning, β is the free-parameter fit to each subject's choice that scales the Q-value. λ is the reward discount parameter. Higher values of w indicate greater recruitment of a MF, while higher values of w indicate more use of a MB learning strategy. Temperature and decay parameters are two memory relevant parameters that describe how noisy a memory is, and how fast a memory is forgotten across time.

Then, the log-likelihood function is calculated for each individual subject using the observed data. Finally, the maximum likelihood estimation technique is used to estimate the parameters of the distribution that best fit the individual subject's data.

To estimate the optimized parameter for each individual subject, we computed the probability of the two states' responses and calculated the log-likelihood of each subject's performance using the SciPy library (Virtanen et al., 2020). Following Nussenbaum et al. (2020), this process is repeated for each subject 10 times in the sample, resulting in estimates of individual subjects' data based on the highest log-likelihood value among 10 optimizations.

5.5 RESULTS

As shown in Figure 5.3, our ACT-R hybrid model is able to replicate the canonical response switch patterns in MF, MB and hybrid RL. Furthermore, our ACT-R model is capable of predicting response time, whereas most RL models fail to do so.

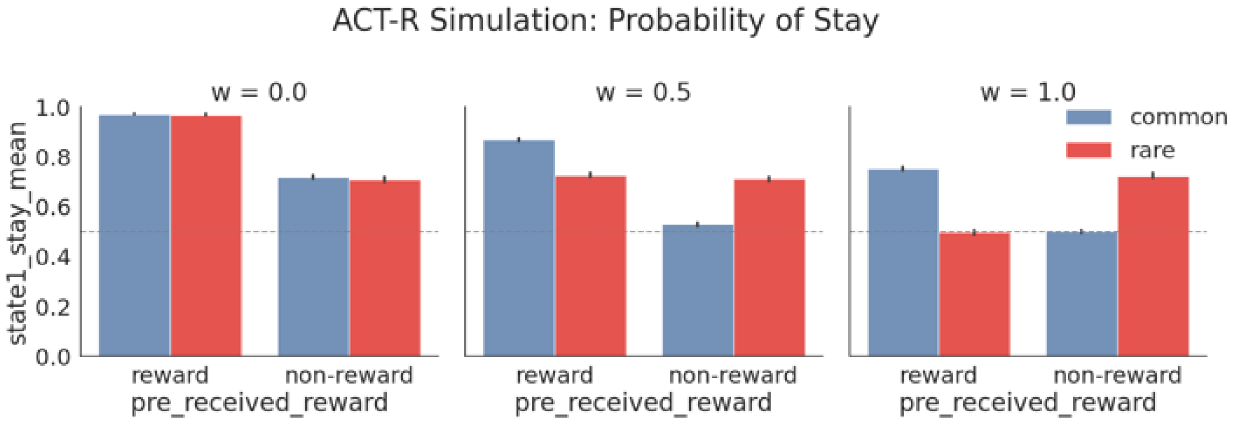

Fig 5.3. ACT-R simulation: The probability of stay by previous reward and transition frequency, with various degrees of MB Learning (w = 0, 0.5, 1)*.* α *= .8,* β *= 5,* λ *= 0.6, temperature = 0.2, decay = 0.5, and each agent simulated 201 trials 100 times.*

When it comes to parameter estimation, our ACT-R model largely agrees with the RL models (See Figure 5.4). We found a significant positive correlation between RL and ACT-R Hybrid optimized parameters (α: $r = 0.24$, $p = 0.0025$; β: $r = 0.27$, $p < 0.001$, λ: $r = 0.31$, $p < 0.001$) and, moreover, the maximum log-likelihood value is positively correlated ($r = 0.86$, $p < 0.001$) between two models.

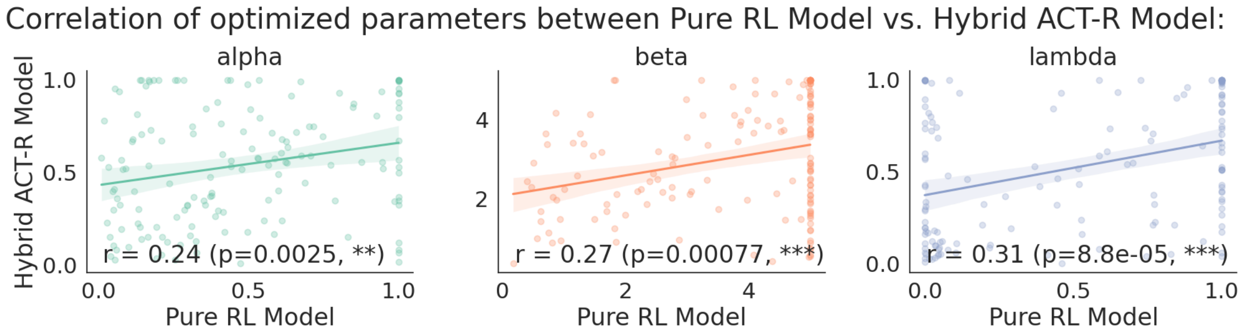

Fig 5.4. The correlation of optimized parameters: alpha, beta and lambda between Pure RL Model and ACT-R Hybrid Model.

We also examined the effect of the decay and temperature parameters of the ACT-R Hybrid Model on the probability of staying. Figure 5.5 demonstrates how the mean probability of staying changes as a function of temperature and decay. As we expect, with lower decay, the agent is better at recalling the state-action association, and in turn, such an accurate estimation of state-action frequency encourages the usage of the MB learning strategy. On the other hand, high decay suggests worse memory recall, leaving agents no choice but to rely on the MF learning strategy.

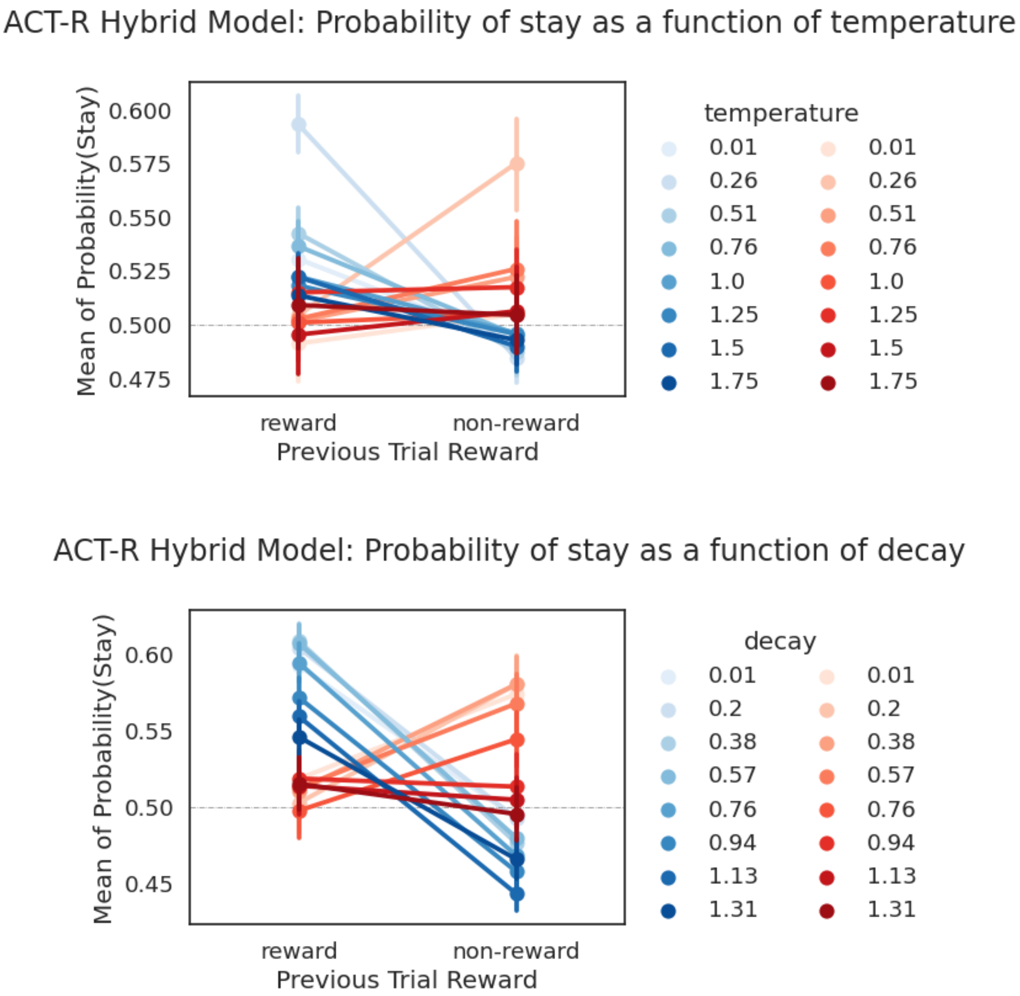

Fig 5.5. The simulated mean probability of stay as a function of two parameters in ACT-R Hybrid Model: temperature (top) and decay (bottom). *The temperature parameter ranges from 0.01 to 1.75, simulating 200 trials 100 times. The decay parameter ranges from 0.01 to 1.4, simulating 200 trials 100 times.*

Effects of Long-Term Memory on Decision Times. According to empirical data from Nussenbaum et al., (2020), it takes longer for participants to respond if a previous trial is rarer than common, as shown in Figure 5.6(Left). Figure 5.6(Right) illustrates the simulated mean response time of common vs. rare trials in the hybrid model. This result is explained by the memory mechanisms in ACT-R, whereby the memory of a more common event has higher activation, which makes it easier and faster to recall. In contrast, this effect is not immediately predicted by the classic, pure RL model.

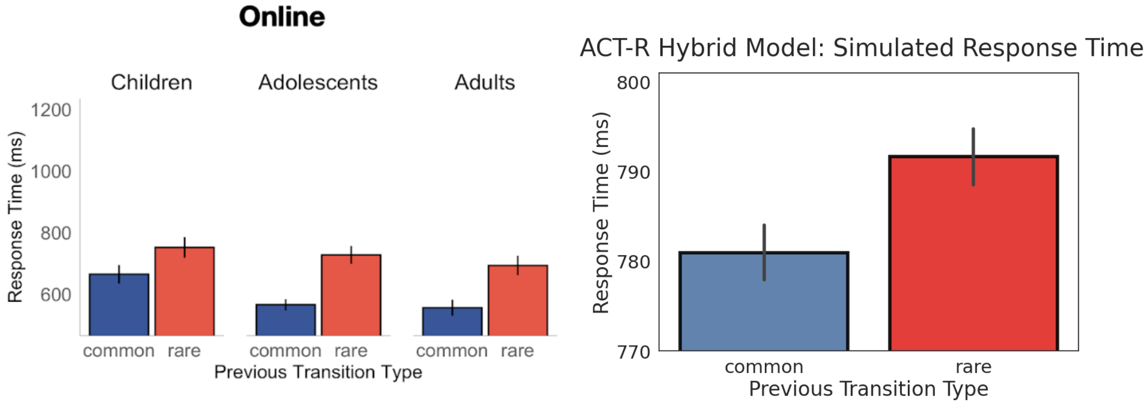

Fig 5.6. Response time (ms) by transition type (common vs. rare) *(Left) This figure is adapted from Nussenbaum et al., (2020) for three age groups. (Right) Simulated response time (ms) of ACT-R Hybrid Model by state transition type.* $\alpha = .8$, $\beta = 5$, $\lambda = 0.6$, *temperature = 0.2, decay = 0.5, lf = 0.15, fixed_cost = 0.1 and simulated 201 trials 100 times. The black line represents the standard error of the mean.*

Interestingly enough, we also delved into how memory parameters, temperature, decay, and latency factor affect agents' ability to estimate transition frequencies in the planning process. ACT-R's memory retrieval process is influenced by two parameters: decay and temperature (or noise). Essentially, high decay equates to memory fading quickly, while high noise signals that memory isn't very reliable. Under normal circumstances, we expect agents to start with random guesses and gradually learn about 70% (common path) and 30% (rare path) transition probabilities through interaction with the environment, as denoted by the dashed gray horizontal lines in Figure 5.7.

Our simulation results suggest that when both decay and temperature parameters both are low (see Figure 5.7 row 1 column 1), indicating that memories never fade and less noisy, and agents have have a tendency to overestimate the transition frequency, regarding the common path as more common (more than 70%), and the rare path more rare (less than 30%). Conversely, when decay is low but temperature is high (see Figure 5.7 row 1 column 3), implying unreliable and non-fading memory, agents tend to underestimate the transition frequency. Specifically, they believe the common path is less common (less than 70%), and the rare path is less rare (more than 30%). And if decay was particularly high, suggesting overall rapid memory fading, the agents tend to stop learning the transition frequency at all.

Fig 5.7. The effect of memory parameters on transition frequency estimation. *The x axis represents the binned 200 trials index (20 bins), and the y axis represents the estimated transition frequency. Gray horizontal lines denote the ground truth transition frequency, 70% for common path and 30% for rare path. Each color represents a state-action transition. For example, (A, 'f', B) means the transition frequency from state A to state B if choosing 'f' option (also called left option).*

<u>Capturing Individual Developmental Differences</u>. The individually-fitted parameters also provide insight into the developmental data observed in Nussenbaum et al. (2020). Specifically, the authors showed that the use of the MB component increases with age. In the pure RL model, this effect can only be explained by altering the weight parameter according to age. When examining the individual parameters of the hybrid models, however, we found a significant negative correlation between the optimized decay rate d (Eq 4) and the age of each subject (r = -0.21, $p < 0.01$), as shown in Figure 5.8, suggesting that younger subjects have a higher decay rate than older subjects. This is in line with the development of cognitive abilities in children, and provides an explanation for why, as people get older, they tend to rely more on declarative memory (MB strategy) in recreating a mental model of the environment.

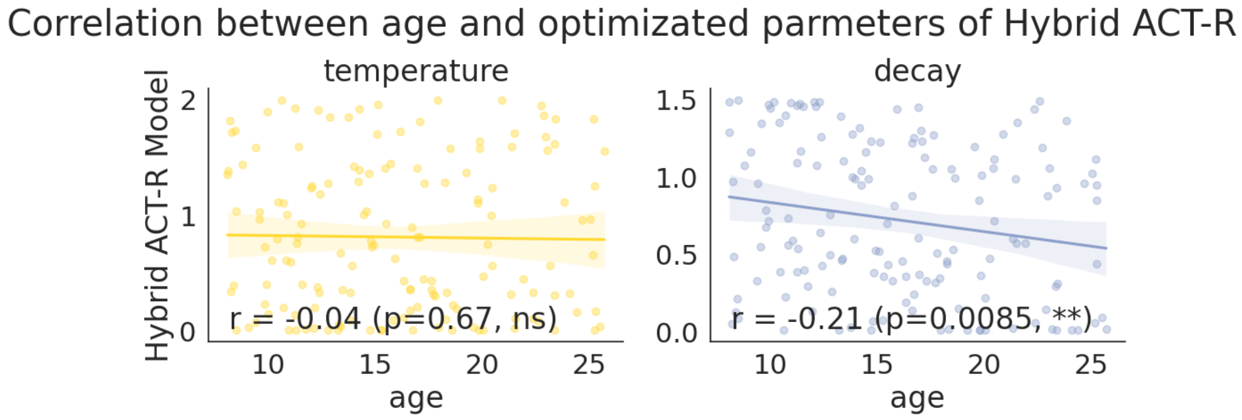

Fig 5.8. The correlation of age and optimized memory relevant parameters from the ACT-R Hybrid Model fit to each subject (N = 151) is shown above. *Each subject was optimized 10 times with random initial parameter seeds. There is a negative correlation between age and the memory decay parameter (r = -0.21, p < 0.01), but no correlation between age and temperature.*

5.6 DISCUSSION

In this paper, we have explored the cognitive underpinnings of decision-making and learning in complex environments, focusing on the interplay between MB and MF reinforcement learning strategies. The declarative framework in ACT-R provides a cognitive representation of how individuals adapt to unstable environments by leveraging memory sampling and incorporating this inherent uncertainty into human decision-making. We propose that MBRL is closely tied to declarative memory, with individuals relying on memory for previous episodic traces when making plans. To test this idea, we implemented an ACT-R model that incorporates declarative memory and procedural Q learning in the decision-making process. Our results provide compelling evidence for the relationship between memory resources and the mixture of MB and MF strategies.

By establishing this connection between our results and the developmental data from Nussenbaum et al. (2020), we provide further evidence for the importance of declarative memory in the decision-making process and highlight the developmental changes that occur in the balance between MB and MF strategies. These findings contribute to a more comprehensive understanding of the cognitive foundations of MBRL and the factors that influence the balance between MB and MF strategies across different age groups. Ultimately, this knowledge may inform the development of age-appropriate interventions and strategies to improve decision-making outcomes throughout the lifespan.

Moreover, our model offers some insight into how memory retention and reliability can affect one's ability to accurately estimate real-world transition probabilities. Such variations in memory parameters may be able to account for why some people are more adept at adapting to changing environments or making predictions based on past experiences. In the real world, memory is never perfect, and how we sample memories may greatly shape how we view the world. Individual characteristics of memory retrieval, such as decay speed or the level of noise, may contribute to individual differences in decision-making process. These findings underline the importance of understanding the interplay of memory factors in decision-making processes, opening doors for future research in individual differences in learning and adaptation strategies.

Lastly, this developed Hybrid model in ACT-R can effectively link back to the EVC model (Chapter 3), creating a cohesive and unified understanding of decision-making processes. Even more, this model transcends the boundaries of task complexity, offering a framework that can interpret the cost associated with decision-making in various cognitive tasks. This spans from simple domains, such as the Simon task, to more complex ones like the probabilistic Markov task. The elegance of this model lies in its ability to quantify and analyze the multifaceted costs that underpin our decisions, including cognitive effort, time, and probabilistic uncertainties. By utilizing the principles of EVC, this model offers a nuanced understanding of how these costs are evaluated and weighed in the process of making a decision, modeling how much effort is needed to deal with memory-based “planning” and when one would turn from “planning” to “habit” thinking. This connective thread between the Markov model and EVC theory facilitates a deeper comprehension of how we adapt to and navigate through complex decision landscapes, demonstrating the critical role that such computational modeling can play in the field of cognitive psychology.

In sum, this study offers a unique perspective on decision-making in dynamic environments, incorporating elements of both the MF and MB approaches and, most importantly, providing a plausible cognitive framework for planning. This synthesis of reinforcement learning principles within the ACT-R framework provides valuable insights into the cognitive mechanisms underlying human decision-making and adaptation. Our work underscores the necessity of considering the contributions of multiple memory systems when investigating the balance between MB and MF approaches to decision-making.

Chapter 6. GENERAL DISCUSSION

6.1 OVERALL SUMMARY

This dissertation aims to advance our knowledge of decision-making with three pieces of computational modeling research. First, in Chapter 3, by extending the EVC framework (Shenhav et al., 2013) to a more detailed model of cognitive processes (i.e., ACT-R), I am able to account for how much effort one is willing to invest towards complex cognitive tasks and model what specific cognitive components are involved in this effort allocation processes. Next, in Chapter 4, two main decision-making processes people use are identified via cognitive modeling, either relying on procedural mechanisms (likely associated with System 1) or declarative memory retrieval mechanisms (likely associated with System 2). By using computational models, I am able to identify the preferential use of these processes in the patterns of decisions made by participants; furthermore, I am able to predict which mechanisms an individual would rely upon from his/her functional connectivity, using machine learning models. Finally, in Chapter 5, I dig into the costly nature of memory in decision-making, and model the transition between easy learning through repetition and more effortful learning that involves memory.

Taken together, these three studies suggest that EVC theory offers a compelling explanation for why people sometimes make decisions that seem "costly", and also why, even in similar circumstances, people's decisions can vary widely. In Chapter 3, I showed that decisions about allocating cognitive resources and effort can be explained as being guided by an evaluation of the cost-benefit trade-off. In essence, individuals are motivated to allocate resources in a way that maximizes the expected benefit (or value) while minimizing the cost, in terms of cognitive effort, time, and other relevant factors. When considering the choice between different strategies, EVC theory suggests that individuals assess the efficacy of each strategy - the potential reward or outcome that each strategy could yield, balanced against the required effort.

For those preferring a seemingly “more costly” choice, their personal decision-making calculations (EVC) may tell them that the benefits of a more demanding strategy are worth the extra effort, making it a better choice. This could be because of a number of factors, such as how hard the decision is perceived to be, how cognitively capable the person making it is, what they expect the result to be, or the situation in which the choice is made. From the perspective of EVC theory, the preference for a declarative strategy can be seen as an adaptive decision-making strategy, because it represents an individualized cost-benefit analysis that leads to the selection of the strategy that is perceived to offer the highest expected value, given the specific circumstances of the decision-maker.

In Study 2 (Chapter 4), I gained insight into the fact that individuals display preferences in their use of either declarative or procedural strategies during decision-making. The fact that a sizable percentage of individuals would consistently prefer using a declarative-memory-based strategy, rather than a procedural strategy, seems to be in apparent contradiction to the EVC framework. After all, memory retrieval (often identified with System 2) is considered more demanding and effortful than simple stimulus-response learning (identified with System 1), so why would any individual use it if it’s not more advantageous? The answer lies in the brain systems that support the two strategies. As my findings show, individuals that prefer to use the memory retrieval strategy also show greater functional connectivity in brain regions associated with memory encoding and retrieval. This, in turns, suggests that, in those individuals that belong to the declarative group, the computations underlying memory retrieval can be carried out as efficiently as those underlying the simple stimulus-response decision-making. Thus, this finding aligns with the EVC framework when we take into account the individual-specific neural costs of decision-making.

Yet, this binary categorization may oversimplify the complex and dynamic nature of how people make decisions. It implies a strict dichotomy that might not reflect the full picture of human decision-making processes. In the real world, the choice between declarative and procedural strategies is not black and white, but rather a matter of degree. It's more accurate to think of these strategies as two ends of a spectrum, with individuals navigating between them based on a variety of factors. Our decision-making behavior isn't just about choosing one strategy over the other but often involves a nuanced balance of both.

This led to the investigation in Study 3 (Chapter 5), which attempted to better understand the underlying reasons that drive this adaptability towards different strategies. Memory, as revealed in Study 3 (Chapter 5), plays a crucial role in a more effortful model-based (MB) planning strategy, which is closely associated with the declarative system. Using memory in decision-making involves retrieving past experiences, comparing options, and predicting future outcomes, which can be cognitively demanding. This "cost" comes in the form of mental effort. This effort is dependent on several factors including the complexity of the task, the individual's cognitive resources, and the time pressure associated with the decision. Here, I delved into the role of memory in decision-making, hoping to shed light on why and when individuals might lean more towards one end of the spectrum or the other. I modeled the transition process between the less effortful procedural strategy to the more cognitively demanding declarative strategy. By doing so, I provided a more detailed and realistic depiction of the decision-making process, capturing the dynamic and adaptable nature of human cognition.

With the insights from the last two studies (Chapter 4 and Chapter 5), it becomes clear that this EVC model (Chapter 3) extends to the choice of decision-making strategies as well. Given that memory use can be costly in terms of cognitive effort, it raises the question: why would around 20% of subjects in Study 2 (Chapter 4) decide to use the declarative strategy, when adopting the declarative strategy seems to be less wise because it requires more cognitive effort? Study 3 may provide a plausible answer: these declarative decision-makers tend to have more cognitive resources available or have a better ability to manage cognitive load for memory retrieval and planning, allowing them to deal with the extra mental effort required for the declarative strategy. It could also be that these individuals perceive the potential benefit of the declarative strategy by outweighing the cognitive effort involved, or the way their brain wired allows them to optimize this high-efficacy declarative-based strategy over procedural-based strategies.

In simpler terms, it's possible that some people choose to use a more effortful "memorize and plan" strategy because they believe it helps them make better decisions, even though it requires more mental work. They might see the benefits of this strategy as being worth the extra effort. It's also likely that the brains of these people are just naturally better at using the "memorize and plan" strategy. They might find it easier or more comfortable to use, and this can make it feel like less effort than it might to others. This could be because of the way their brain is wired from birth, or because of their experiences and how they've learned to think and solve problems. So the reason some people prefer the "memorize and plan" strategy could be a mix of believing in its benefits and being naturally good at it. In short, individuals preferring a declarative or “memorize and plan” strategy, despite its higher cognitive effort cost, are likely doing so because their internal EVC calculation justifies this choice. They estimate the value of outcomes from using this strategy to outweigh the cognitive costs, resulting in a higher EVC. Understanding these reasons can help us learn more about how different people make decisions.

Even though some strategies might appear less demanding on the surface, it doesn’t necessarily make them the universally preferred choice. Individuals’ unique EVC calculations take into account their cognitive architecture, decision-making preferences, and the perceived value of outcomes. These factors can tilt their choice towards more effortful strategies, illustrating how each person's decision-making process can vary greatly. This understanding further underscores the importance of considering cognitive effort and individual differences in studying decision-making.

In essence, the decision to use whichever strategy can be seen as a balance between the perceived benefit and the cognitive effort involved. Our computational research provides a better understanding of this balance and how it shapes our decision-making processes. While there is still much to learn about what other factors might influence an individual’s preference for different decision-making strategies, these studies collectively suggest that memory, willingness to invest effort, and cognitive resources play significant roles in this complex process.

6.2 FUTURE DIRECTIONS

Based on what was learned from the three studies in this dissertation, there are a number of interesting directions for future research in cognitive psychology. Future research could explore the heuristics and biases that lead to “irrational” decisions. For instance, why do individuals consistently make decisions that violate principles of rationality, like choosing short-term gratification over long-term benefits? Are there certain conditions under which these biases are more likely to manifest? This dissertation improves our understanding of the cognitive and neurological processes involved in decision-making; however, the decision-making environment is far less complicated compared to real life.

Investigating how decision-making varies across the lifespan could offer insights into whether these decisions are linked to cognitive development or cognitive decline. In Chapter 5, I provided one possible explanation of how developmental change in memory, especially the memory decay, may affect one’s reliance on one strategy over another. Further research could include even older individuals and see how decision-making would change when, in an elderly population, memory function begins to decline. Would the aged participants look again like the very young ones?

It's widely recognized that emotions can significantly influence decision-making, often leading to decisions that may appear more “costly” from a purely rational perspective. More research could delve into the interplay between emotions and cognition in decision-making, exploring questions like how does mood influence risk perception and decision-making? Can emotional regulation strategies help improve decision quality? Computational research is important in advancing understanding of how external factors and stressors affect decision-making. For example, how do time pressure, emotional arousal, or peer influence shape decisions, especially as it affects the neural computation? ? Is there any way to quantify time pressure or emotional stress in EVC evaluation? Do these factors amplify tendencies to make more effortful decisions? If so, how?

Addressing these questions will significantly contribute to our understanding of how people make decisions, and why they make decisions so differently, and finally will enable the development of strategies to promote better decision-making in various contexts.

6.3 SIGNIFICANCE

The work presented in this dissertation holds profound significance. First, it elucidates the nuanced processes underlying decision-making, revealing how cognitive control, memory, and individual differences intersect in this critical aspect of human cognition. Our investigations span simple to complex tasks, providing comprehensive insights into the decision-making field in cognitive science. Second, the use of cognitive modeling, especially the exploration and implementation of the EVC Model within the ACT-R framework, has given us a very important new tool for analysis. These models allow us to not only quantitatively dissect decision-making processes but also test our hypotheses in a controlled, repeatable way. The predictive power of these models can guide future empirical research and eventually inform real-world applications, such as designing more human-like intellectual artificial intelligence or improving educational strategies. Moreover, the unique focus on individual differences and decision-making strategies, particularly the transition between declarative and procedural strategies, contributes a valuable perspective to the field. It demonstrates the diversity in decision-making and emphasizes the need to take this variability into account in computational modeling frameworks. Lastly, this dissertation work sets a direction for future research, posing key questions and potential avenues of inquiry. In an era where decision-making is increasingly complicated by the onslaught of information and options, understanding why and how people make certain decisions – logical or otherwise – becomes increasingly pertinent.

Through three studies, I establish a unified understanding of the cost inherent in decision-making by synergizing the insights derived from the Expected Value of Control (EVC) theory. The primary takeaway from this research is the recognition that every decision I make is contingent upon an inherent cost, be it cognitive, time-bound, or resource-intensive. Our newly developed hybrid model of ACT-R (Chapter 5) provides an innovative platform to dissect these costs, enabling a deeper understanding of what truly drives our decision-making processes. Furthermore, this model offers a cognitively plausible methodology to examine probabilistic decision-making. It mimics the human cognitive system's ability to process information and make decisions under uncertain conditions, thereby bringing us closer to replicating human-like decision-making in artificial systems. In summary, the nexus between our proposed hybrid model and EVC theory offers a unique and powerful framework for understanding the cost of decision-making, underscoring the critical role of computational modeling in cognitive psychology.

BIBLIOGRAPHY

Adams, E., & Simon, H. A. (1962). Models of Man, Social and Rational: Mathematical essays on rational human behavior in a social setting. *The Journal of Philosophy*, *59*(7), 241–260. https://doi.org/10.2307/2023734

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-Motivated learning: Mesolimbic activation precedes memory formation. *Neuron*, *50*(3), 507–517. https://doi.org/10.1016/j.neuron.2006.03.036

Anderson, J. R. (2007). How can the human mind occur? In *How Can the Human Mind Occur in the Physical Universe?* (pp. 237–248). Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195324259.003.0006

Anderson, J. R. (2009). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R. (2013). *Cognitive skills and their acquisition* (p. 384). Psychology Press.

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought* (p. 504). Psychology Press.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., & Van Essen, D. C. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1–3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berridge, K. C., & Valenstein, E. S. (1991). What psychological process mediates feeding evoked by electrical stimulation of the lateral hypothalamus? *Behavioral Neuroscience*, *105*(1), 3–14. https://doi.org/10.1037/0735-7044.105.1.3

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology*, *66*(1), 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 16–27. https://doi.org/10.3758/cabn.9.1.16

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., Adcock, R. A., Barch, D. M., Botvinick, M. M., Carver, C. S., Cools, R., Custers, R., Dickinson, A., Dweck, C. S., Fishbach, A., Gollwitzer, P. M., Hess, T. M., Isaacowitz, D. M., Mather, M., … Somerville, L. H. (2014). Mechanisms of motivation–cognition interaction: Challenges and opportunities. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 443–472. https://doi.org/10.3758/s13415-014-0300-0

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222. https://doi.org/10.1016/s1364-6613(00)01483-2

Chelian, S. E., Paik, J., Pirolli, P., Lebiere, C., & Bhattacharyya, R. (2015, August). *Reinforcement learning and instance-based learning approaches to modeling human* *decision making in a prognostic foraging task*. 2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). http://dx.doi.org/10.1109/devlrn.2015.7346127

Conway-Smith, B., & West, R. L. (2023). Clarifying System 1 & 2 through the Common Model of Cognition. arXiv preprint arXiv:2305.10654.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W. (2012). AFNI: What a long strange trip it's been. *NeuroImage*, *62*(2), 743–747. https://doi.org/10.1016/j.neuroimage.2011.08.056

D, M. (2019). *A lightweight Python implementation of a subset of the ACT-R cognitive architecture's Declarative Memory.* https://pypi.org/project/pyactup

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, *84*(6), 3072–3077. https://doi.org/10.1152/jn.2000.84.6.3072

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of Learning and Memory*, *117*, 4–13. https://doi.org/10.1016/j.nlm.2014.04.014

Esposito, F., Otto, T., Zijlstra, F. R. H., & Goebel, R. (2014). Spatially distributed effects of mental exhaustion on resting-state FMRI networks. *PLoS ONE*, *9*(4), e94222. https://doi.org/10.1371/journal.pone.0094222

Evans, J. St. B. T. (2008). Dual-Processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process theories of higher cognition. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences*, *103*(31), 11778–11783. https://doi.org/10.1073/pnas.0602659103

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1). https://doi.org/10.18637/jss.v033.i01

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience*, *21*(23), 9430–9437. https://doi.org/10.1523/jneurosci.21-23-09430.2001

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 113–131. https://doi.org/10.1080/14639220210159717

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., Hertwig, R., & Pachur, T. (2015). *Heuristics: The foundations of adaptive behavior* (p. 0). Oxford University Press, USA.

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2008). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, *19*(2), 483–495. https://doi.org/10.1093/cercor/bhn098

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, *118*(4), 523–551. https://doi.org/10.1037/a0024558

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635. https://doi.org/10.1207/s15516709cog2704_2

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. W. (2018). *Motivation and cognitive control in depression*. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/500561

Haile, T., Prat, C., & Stocco, A. (2020). One Size Doesn't Fit All: Idiographic Computational Models Reveal Individual Differences in Learning and Meta-Learning Strategies. *Proceedings of the 18th International Conference on Cognitive Modeling.* , 75–81.

Halverson, T., Myers, C. W., Gearhart, J. M., Linakis, M. W., & Gunzelmann, G. (2022). Physiocognitive modeling: Explaining the effects of caffeine on fatigue. *Topics in Cognitive Science*, *14*(4), 860–872. https://doi.org/10.1111/tops.12615

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology* (pp. 139–183). Elsevier. http://dx.doi.org/10.1016/s0166-4115(08)62386-9

Hartmann, M. N., Hager, O. M., Reimann, A. V., Chumbley, J. R., Kirschner, M., Seifritz, E., Tobler, P. N., & Kaiser, S. (2014). Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophrenia Bulletin*, *41*(2), 503–512. https://doi.org/10.1093/schbul/sbu102

Herlambang, M. B., Cnossen, F., & Taatgen, N. A. (2021). The effects of intrinsic motivation on mental fatigue. *PLOS ONE*, *16*(1), e0243754. https://doi.org/10.1371/journal.pone.0243754

Herlambang, M. B., Taatgen, N. A., & Cnossen, F. (2021). Modeling motivation using goal competition in mental fatigue studies. *Journal of Mathematical Psychology*, *102*, 102540. https://doi.org/10.1016/j.jmp.2021.102540

Hill, N. M., & Schneider, W. (2006). Brain changes in the development of expertise: Neuroanatomical and neurophysiological evidence about skill-based adaptations. In *The Cambridge Handbook of Expertise and Expert Performance* (pp. 653–682). Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511816796.037

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), 497–498. https://doi.org/10.1038/nn1238

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. https://doi.org/10.1016/j.tics.2011.12.008

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences*, *107*(19), 8871–8876. https://doi.org/10.1073/pnas.1002007107

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kennerley, S. W., Behrens, T. E. J., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, *14*(12), 1581–1589. https://doi.org/10.1038/nn.2961

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. https://doi.org/10.1371/journal.pcbi.1002055

Knoch, D., & Fehr, E. (2007). Resisting the power of temptations: The right prefrontal cortex and self-control. *Annals of the New York Academy of Sciences*, *1104*(1), 123–134. https://doi.org/10.1196/annals.1390.004

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106–120. https://doi.org/10.1101/lm.1.2.106

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*(1), 131–141. https://doi.org/10.1037/a0031048

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*(12), 899–908. https://doi.org/10.1038/s41562-018-0401-9

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLOS Computational Biology*, *12*(8), e1005090. https://doi.org/10.1371/journal.pcbi.1005090

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kotseruba, I., & Tsotsos, J. K. (2018). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53*(1), 17–94. https://doi.org/10.1007/s10462-018-9646-y

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, *35*(1), 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512

Lovett, M. C. (2005). A strategy-based interpretation of stroop. *Cognitive Science*, *29*(3), 493–524. https://doi.org/10.1207/s15516709cog0000_24

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, *36*(2), 265–284. https://doi.org/10.1016/s0896-6273(02)00974-1

Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, *34*(2–3), 205–236. https://doi.org/10.1016/0301-0511(92)90016-n

Musslick, S., Shenhav, A., Botvinick, M., & Cohen, J. (n.d.). *A computational model of control allocation based on the expected value of control*. In The 2nd multidisciplinary conference on reinforcement learning and decision making.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. https://doi.org/10.1016/s0022-2496(02)00028-7

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368. https://doi.org/10.1002/1098-1063(2000)10:4<352::aid-hipo2>3.0.co;2-d

Nagashima, K., Morita, J., & Takeuchi, Y. (2021). Modeling intrinsic motivation in ACT-R. *Transactions of the Japanese Society for Artificial Intelligence*, *36*(5), AG21-E_1-13. https://doi.org/10.1527/tjsai.36-5_ag21-e

Nee, D. E., & D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife*, *5*. https://doi.org/10.7554/elife.12112

Newell, A., & Rosenbloom, P. S. (1982). *Mechanisms of skill acquisition and the law of practice* (p. 52).

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. https://doi.org/10.1016/j.jmp.2008.12.005

Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra: Psychology*, *6*(1). https://doi.org/10.1525/collabra.17213

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, *47*(6), 419–427. https://doi.org/10.1037/h0058775

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning. *Psychological Science*, *24*(5), 751–761. https://doi.org/10.1177/0956797612463080

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), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge University Press. http://dx.doi.org/10.1017/cbo9781139173933

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546–550. https://doi.org/10.1038/35107080

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251. https://doi.org/10.1016/s0028-3932(02)00157-4

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, *7*(2), 155–162. https://doi.org/10.1093/comjnl/7.2.155

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, *37*(4), 1083–1090. https://doi.org/10.1016/j.neuroimage.2007.02.041

Rescorla, R. A., & Wagner, A. R. (1972, January 1). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. https://www.researchgate.net/publication/233820243_A_theory_of_Pavlovian_conditioning_Variations_in_the_effectiveness_of_reinforcement_and_nonreinforcement

Robinson, T. E., & Berridge, K. C. (2003). Addiction. *Annual Review of Psychology*, *54*(1), 25–53. https://doi.org/10.1146/annurev.psych.54.101601.145237

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997 Mar 14;275(5306):1593-9. doi: 10.1126/science.275.5306.1593. PMID: 9054347.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, *8*(1), 305–321. https://doi.org/10.1111/tops.12183

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *The Journal of Neuroscience*, *31*(12), 4407–4420. https://doi.org/10.1523/jneurosci.3335-10.2011

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

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. https://doi.org/10.1016/j.neuron.2013.07.007

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*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*(1), 146–148. https://doi.org/10.1037/0021-9010.72.1.146

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99. https://doi.org/10.2307/1884852

Simon, H. A. (1957). Models of Man, Social and rational. Mathematical essays on rational human behavior in a social setting. Herbert A. Simon. Wiley, New York; Chapman & Hall, London, 1957. 287 pp. $5. *Science*, *126*(3263), 241–260. https://doi.org/10.1126/science.126.3263.85.a

Smith, B. M., Thomasson, M., Yang, Y. C., Sibert, C., & Stocco, A. (2021). When Fear Shrinks the Brain: A computational model of the effects of posttraumatic stress on hippocampal volume. *Topics in Cognitive Science*, *13*(3), 499–514. https://doi.org/10.1111/tops.12537

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, *80*(4), 914–919. https://doi.org/10.1016/j.neuron.2013.08.009

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386. https://doi.org/10.1126/science.1896849

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Stocco, A. (2017). A biologically plausible action selection system for cognitive architectures: Implications of basal ganglia anatomy for learning and decision‐making models. *Cognitive Science*, *42*(2), 457–490. https://doi.org/10.1111/cogs.12506

Stocco, A., Murray, N. L., Yamasaki, B. L., Renno, T. J., Nguyen, J., & Prat, C. S. (2017). Individual differences in the Simon effect are underpinned by differences in the competitive dynamics in the basal ganglia: An experimental verification and a computational model. *Cognition*, *164*, 31–45. https://doi.org/10.1016/j.cognition.2017.03.001

Stocco, A., Prat, C. S., & Graham, L. K. (2021). Individual differences in reward‐based learning predict fluid reasoning abilities. *Cognitive Science*, *45*(2). https://doi.org/10.1111/cogs.12941

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. https://doi.org/10.1007/bf00115009

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bulletin*, *2*(4), 160–163. https://doi.org/10.1145/122344.122377

Tanji, J., & Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiological Reviews*, *88*(1), 37–57. https://doi.org/10.1152/physrev.00014.2007

Taylor, J. A., & Ivry, R. B. (2014). Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. In *Progress in Brain Research* (pp. 217–253). Elsevier. http://dx.doi.org/10.1016/b978-0-444-63356-9.00009-1

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, *398*(6729), 704–708. https://doi.org/10.1038/19525

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6), 2005-2015.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

van Vugt, M. K., & van der Velde, M. (2018). How does rumination impact cognition? A first mechanistic model. *Topics in Cognitive Science*, *10*(1), 175–191. https://doi.org/10.1111/tops.12318

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). Author Correction: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 352–352. https://doi.org/10.1038/s41592-020-0772-5

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van

Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2014). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*(3), 345–352. https://doi.org/10.1038/mp.2014.44

Weissengruber, S., Lee, S. W., O'Doherty, J. P., & Ruff, C. C. (2019). Neurostimulation reveals context-dependent arbitration between model-based and model-free reinforcement learning. *Cerebral Cortex*, *29*(11), 4850–4862. https://doi.org/10.1093/cercor/bhz019

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE*, *8*(7), e68210. https://doi.org/10.1371/journal.pone.0068210

Wilson, R. C., & Collins, A. (2019). *Ten simple rules for the computational modeling of behavioral data*. Center for Open Science. http://dx.doi.org/10.31234/osf.io/46mbn

Yang, Y. C., Karmol, A. M., & Stocco, A. (2021). Core cognitive mechanisms underlying syntactic priming: A comparison of three alternative models. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.662345

Yang, Y. C., & Stocco, A. (2023). Allocating Mental Effort in Cognitive Tasks: A Model of Motivation in the ACT-R Cognitive Architecture. *Under Review*.

Yang, Y., Sibert, C. L., & Stocco, A. (2023). *Reliance on episodic vs. procedural systems in decision-making depends on individual differences in their relative neural efficiency*. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/2023.01.10.523458

Yang, Y., Stocco, A., & Sibert, C. (2021, July). Individual differences in decision making strategies can be predicted by resting-state functional connectivity. Paper presented at Virtual MathPsych/ICCM 2021. Via mathpsych.org/presentation/631.

Yang, Y., & Stocco, A. (2023, July). *The Cognitive Substrates of Model-Based Learning: An Integrative Declarative-Procedural Model.* Paper presented at Virtual MathPsych/ICCM 2022. Via mathpsych.org/presentation/1279.

Yang, Y., & Stocco, A. (2022, July). *A model of motivation and effort allocation in the ACT-R cognitive architecture.* Paper presented at Virtual MathPsych/ICCM 2022. Via mathpsych.org/presentation/852.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295x.111.4.931

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Zhou, P., Sense, F., van Rijn, H., & Stocco, A. (2021). Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. *Cognition*, *212*, 104660. https://doi.org/10.1016/j.cognition.2021.104660