Understanding Individual Differences in Learning Strategies, Cognitive Characteristics and Task Demands: One-Size Does Not Fit All but Tailoring is Difficult

Theodros Haile

A dissertation

submitted in partial fulfillment of the

requirements for degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Chantel S. Prat, Chair

Andrea Stocco

Susan Joslyn

Ione Fine

Program Authorized to Offer Degree:

Department of Psychology

University of Washington

Abstract

Understanding Individual Differences in Learning Strategies, Cognitive Characteristics and Task Demands: One-Size Does Not Fit All but Tailoring is Difficult

Theodros Haile

Chair of the Supervisory Committee:

Chantel S. Prat

Department of Psychology

This dissertation examined the fundamental aspects of skill learning, exploring three key facets: the use of different strategies by learners, the stability of these strategies over time and with varying task requirements, and relating latent cognitive characteristics to learning success or strategy. Idiographic computational models and parameter estimation based on the ACT-R cognitive architecture were used to identify strategies and estimate latent cognitive characteristics. Results showed that participants fit diverse memory strategies, with idiographic modeling proving to be crucial in uncovering these differences. However, a declarative long-term memory (LTM) strategy best described most participants. Taking into account that strategies might change across time or in response to changing task demands, a second experiment investigated individual dynamics and learner-task interactions. Here, the learning task was split into 2-time epochs and fit individually, to test if different models explained

behavior at different time points. For the simple stimulus-response task that was used, most participants were best fit by the LTM strategy throughout the task but increasing task difficulty did not effect consistent changes in strategies. Lastly, a third experiment sought to robustly estimate model parameters that were historically related to cognitive characteristics vital to learning: memory decay rate and working memory capacity. The goal after estimation was to test how predictive these are of learning outcomes in multiple cognitive tasks and strategies. The study concluded that parameter values serve as reliable measures of individual cognitive characteristics only within specific models or contexts and not across tasks.

Table of Contents

Chapter 1: Introduction

Complex skills are vital for individuals to thrive in their environments, and the ability to learn them drives success in a number of situations. But our current ways of teaching new skills and assessing mastery has a one-size fits all approach that disadvantages not only individuals, but also entire communities. Motivated by the findings of skill learning studies (e.g., Prat et al., 2020; Shute, 1991; Tenison et al., 2016), this current doctoral program of research aims to capture the different ways individuals acquire new skills and relate them to idiosyncratic cognitive characteristics.

The hypotheses that motivate this work were developed as a way of elucidating research showing that individual learners, given the same amount of time and resources, do so at different rates (Yamasiki et al, 2018; Prat et al; 2019; Prat et al, 2020). In Prat et al., (2020), participants learned Python programming language over 10 weeks using an online tool and demonstrated that some participants far exceeded others in the number of course levels they reached in those 10 weeks. The study also showed that cognitive characteristics like working memory (measured through span tasks, Oswald et al., 2015), general intelligence (measured through the Raven's Advanced Matrices, Arthur & Day, 1994) and language aptitude (measured through the Modern Language Aptitude Test, Carroll & Sapon, 1959) were significant predictors of rate of learning and programming accuracy. This suggests that a rich set of individual characteristics, like cognitive capacities and previous experiences with learning and language use, affect how we approach learning new skills (Shute, 1991, is another example of such a study). However, the majority of complex skill learning studies simply average over these individual differences. Thus, a major motivating factor of the current program of research is to explore the way individual characteristics shape the different paths people take when acquiring complex skills.

This introduction aims to ground our approach in the existing literature that influenced it. This overview establishes the following distinctions: first, that there are multiple memory systems with different characteristics; second, that skill learning is a multi-stage process that differs with increasing experience and expertise; third, different tasks might require engagement with a specific memory mechanism for optimal performance; and finally, that individuals might rely on distinct, idiosyncratic combinations of memory mechanisms to accomplish learning in early stages.

Brief Introduction to The Multiple Memory Systems View

Memory, both structurally (e.g., Hill & Schneider, 2006) and functionally, consists of multiple interconnected systems (Squire, 2004; Poldrack et al., 2001). This is evidenced by a large number of discipline-spanning studies, from clinical patient studies (e.g., Knowlton et al., 1994) to computational models (e.g., Anderson, 1982; see Keren & Schul, 2009 for arguments against two-system views of memory). These systems are often broadly categorized into explicit and implicit memory categories. The difference between these two types of memory is largely the conscious access, or lack thereof, to information (e.g., Eichenbaum & Cohen, 2001; Squire, 2004), and therefore is synonymous with the declarative-procedural way of categorizing memory systems. While these two broad categories contain all the ways memory systems acquire associations about the world, like, episodic memory, classical conditioning or priming, the focus of this research is only on declarative semantic memories and procedural or skill memories, described briefly next.

Explicit, or declarative, memory is the storage of facts or associations between concepts that one can *declare* (Anderson, 2000), like “green light means go” or “Addis Ababa is the capital city of Ethiopia”. These types of memories are associated with the hippocampus (especially in the case of recently acquired memories) and cortex (especially in the case of consolidated, long-term or early memories; Squire, 2004; Poldrack et al., 2001).

In contrast, implicit or procedural memories are *not declarable*, and while best exemplified by motor skills like riding a bicycle, are stores for more complex sets of associations that accumulate over time. While these two types of memory systems seem vastly different, Anderson (2007), for example, argues that all of memory are associations. From one of the examples above, the chunk of memory that holds “green light means go”, simply maintains a meaningful association among the concepts “green”, “light” and “go”. The “riding a bicycle” procedure, just as simply, represents a vast set of associations between multi-sensory inputs (curve of the road, bicycle tilt etc.), and motor movements, and can possibly be coded by a set of “If ... then” procedures (Anderson, 1982; Anderson, 2007). By suggesting these two examples, I am hoping to surface our understanding of why multiple types of memory might have been evolutionarily beneficial and how it aligns with skill learning theories.

Declarative information like “Green light means go” is much simpler than the complex set of procedures that are needed to successfully navigate on and ride a bike. We can guess at its simplicity by judging the effort and time it takes to recall and say the short phrase, which must be much less time and effort compared to the complexity of operating a bicycle, since the latter involves many such processes. But we can watch any experienced cyclist and notice that all the decisions and behavioral executions they make, despite the skill’s apparent complexity, seem effortless and seemingly take no time. This is true only once the cyclist has had sufficient practice. Novice cyclists might spend much more *effort* simply to balance on the bicycle let alone concurrently navigate, avoid obstacles, and change gears.

Executing any set of new procedures, like those described above requires retrieval from memory and maintenance in working memory, which is capacity limited and requires effort. And this is closely tied with explicit memories - to use and share declarative knowledge, it must be retrieved from long-term memory and maintained in working memory. Using explicit memories allows for fast learning and updating, flexibility in searching for vital information, connecting to new information, formulating a decision or even many other alternative decisions (Anderson, 2007). But there is an upper bound on how much and how fast we can make use of declarative memory, given caps on working memory, which in certain situations might not be ideal. Enter procedural memory, which after much practice, allows for rapid responses and are virtually not capped on the number of procedures that can be executed. To reiterate a point above, this is not limited to motor skills. Anderson, (1982) for example, describes early studies where highly skilled office assistants could efficiently proof-read text while answering calls on the time. The caveat for procedural skills here being that they are inflexible and once started, are difficult to interrupt and change.

In summary, declarative memory use is resource limited but flexible, and procedural memories are unlimited and fast but inflexible, and are acquired after much practice. At the most basic level, we live in an environment where each passing moment might call for either flexibility and slow problem solving, or reactive fluent responses, which might explain the evolutionary utility of memory and skill (Hikosaka et al., 2013).

The Multi-stage Theory of Skill Learning

Acquisition of expertise in any skill is typically described as being a multi-stage process. Popularized by Fitts and Posner (1967), a two- (e.g., Sloman, 1996), three- (e.g., Anderson,

1982) or more (Dreyfus & Dreyfus, 1980) stage of skill acquisition, each with their distinct set of properties and behavioral indicators, is an enduring part of the skill learning literature.

Briefly, early ‘novice’ stages of learning are characterized by the slow, laborious application of a set of procedures to reach an outcome in a given problem space (Anderson, 1982; Tenison et al., 2016). Here, the learner relies on working memory representations of the procedures, which requires constant maintenance, and is either activated by referring to instruction material or by retrieving it from long-term memory (Anderson, 1982). Note that this stage is almost always described as being supported by declarative memory and the capacity limited attention and working memory resources (which are important bottlenecks to performance).

There is often an intermediate stage, where some of this laborious mapping of inputs to outputs, with practice approaches automaticity and relies *less* on the capacity of limited resources like working memory because declarative memory retrieval requests for some of the procedures are no longer necessary (Tenison et al., 2016; Fits & Posner, 1967). In the final ‘expert’ stage, behavior is marked by fluent, rapid, and highly accurate responses. Here, it is understood that the learner now has replaced the declarative maps from inputs to behavioral outputs, with automatic, instinctive, procedural, stimulus-motor response maps that are difficult to interrupt and almost entirely circumvent working memory resources (e.g., Shiffrin & Schneider, 1977; Anderson et al., 1997). The skill at this late stage is thought to be entirely represented by procedural memory and becomes encapsulated from declarative access (i.e., experts can demonstrate skill but might struggle with describing how they arrived at some or all of the solutions).

In the current study, we hypothesize that individual differences in how participants engage with learning *in the early stages*, dictates how well and how quickly they arrive at the late stages of expertise. It should be noted that this is a fairly uncommon perspective on skill acquisition, and one that has not been well investigated, especially with computational models.

Memory Mechanism and Task Relationship

Given the sections above, there are two points that deserve attention to make way for how individual differences might arise: memory system capacity limitations and task requirements and are best explained by laboratory category learning experiments. The examples above, for declarative type memories and procedural type memories were deliberate in how the complexity of the task fit the memory system. Category learning tasks often have conditions that vary the complexity of the features that signify category membership (e.g., one feature (easy) or 2 or more features (hard), as in Ashby & Crossley, 2010 or DeCaro et al., 2008). These tasks might also have features that are only sometimes related to the category label according to a pre-set probability. Note the importance of this: associations that have a low predictive probability are hard to learn, and probabilities that are at chance cannot be learned at all since this violates the very definition of associations (e.g., Weather Prediction Task, Knowlton et al., 1996).

In complex categorization tasks, given that access to and manipulation of explicit memories are capacity limited because they require active maintenance in working memory, and tend to decay rapidly, relying on working memory in these situations might lead to more errors (e.g., DeCaro et al., 2008). For example, it will be easy to learn two categories of objects that are distinguished by one feature, like color, but it will be hard if there are more than two features, or two categories, and, if the features are related to the categories at low probabilities. It can also be more difficult to learn categories if the features are not readily perceptible (e.g., Frank et al., 2004). If, however, we engage with these tasks procedurally, there are no limitations to the number of features and categories that we can maintain, and given practice, will lead to higher accuracy (Collins, 2018; Frank et al., 2004) . In summary, simpler tasks are best suited for declarative memories because these can be learned quickly using explicit associations, and more complex situations, while slower, are best learned using procedural memory, suggesting that some tasks are best learned with one or the other memory systems.

Multiple studies with amnesiac and Parkinson's disease patients demonstrate this division. Influential work by Knowlton and others have shown that amnesic patients, who have impaired declarative memory are able to learn associations between probabilistic cues and their categories (Probabilistic Classification task, Knowlton et al., 1994; Knowlton et al., 1996), while patients with Parkinson's, who have impaired procedural and habit learning perform poorly on the probabilistic classification task (Knowlton et al., 1996), a situation often known as double dissociation. In contrast, Parkinson's disease patients perform just as well as controls on declarative memory tasks, while amnesic patients perform much more poorly (Knowlton et al., 1996). Poldrack et al., (2001), also demonstrated similar dissociation in healthy participants by comparing brain activity in response to a deterministic, simple association task (showed increased activity in medial-temporal lobe (MTL) versus the same probabilistic task as above (increased activity in caudate nucleus of the basal ganglia).

While these results are compelling none have shown definitive evidence that singularly individual systems or brain regions are responsible for representing and guiding behavior in these categorization tasks. For instance, in Knowlton et al., (1996), amnesic patients are never as good as healthy controls, and are outperformed by the latter near the end of the task. This suggests that declarative and procedural systems, at least, process information in parallel, and might collaborate (Collins et al., 2017). Others, on the other hand, suggest that this interaction might be inhibitory (Poldrack et al., 2001; Poldrack & Packard, 2003) and complex tasks are rarely supported by individual memory systems (Ashby & Crossley, 2010).

One additional point of view that influenced our goals must be discussed here. Taken together, the above observations suggest that some tasks may be better suited to learning by one memory system and not the other, but they almost always interact in intricate ways, especially in complex task situations. However, it should be noted that individuals likely adopt different ways of learning complex tasks. This is perhaps because some level of success can be achieved by utilizing different combinations of declarative and procedural memory systems, what we call a *strategy*. These strategies are likely to be different among people. For instance, as examined by Gluck, (2002), participants might rely more on explicit memory and learn the probabilistic classification task by just focusing on the cues that predict category membership with high probability. Other studies also suggest that contextual information affects behavior, where people are more likely to use available declarative information rather than procedural trial-and-error learning (Delgado et al., 2005; Delgado and Dickerson, 2012), or when the task does not strain working memory resources (Collins et al., 2017).

In the current study, given the diversity of ways memory mechanisms can be used to engage with a task, we hypothesize that individual differences in skill learning arise because learners might be deploying different strategies for a given task. It should be noted here that this view and examination of individual strategy use is uncommon. We further seek to answer if learning behavior might be explainable by *differences* in capacity limitations, explored in the next section.

Relating Individual Difference in Cognitive Characteristics to Strategies

Our hypothesis that strategies, and therefore learning behavior, might be driven by individual differences in cognitive characteristics arose from psychometric (e.g., Yamasaki et al., 2018; Prat et al., 2015) experiments and studies of model parameters (e.g., Lovett et al, 2000; Xu & Stocco, 2021), both of which also include the widely studied effects of working memory capacity.

To start working memory capacity, has been shown to be related to virtually all aspects of cognitive and goal-directed behavior (Klingberg, 2010; Goldman-Rakic, 1995) from the very simple, like control of eye movements, where participants are instructed to not make eye-movements (Unsworth et al., 2004), to the complex, like reading comprehension (Prat et al., 2015) reasoning (Baddeley & Hitch, 2007), measures of general intelligence (e.g., Ackerman et al., 2005), and language learning (Just and Carpenter, 1992; Miyake & Friedman, 1998). In most cases, higher working memory capacity is associated with better information processing (e.g., Long & Prat, 2002; Just & Carpenter, 1992) but sometimes, higher working memory capacity might lead to lower performance (DeCaro et al., 2008) perhaps due to individual strategy differences. In DeCaro et al. (2008) for instance, high-capacity individuals likely relied too heavily on working memory to learn complex categories while lower capacity individuals, likely taking a procedural approach, performed better. This suggests an interaction between task demand, cognitive characteristics, and strategy use, which we sought to address with formal computational models in the current study.

In addition to working memory capacity, other features of cognition also might affect individual learning strategies. These features are more apparent when considering learning behavior through computational models of cognition. Collins (2018), one of the studies that had a large impact on hypothesis development, presented a model that contained only working memory and procedural memory (computationally represented by reinforcement learning algorithms) as components. While working memory presents a substantial bottleneck to declarative memory use, there are other well documented features of declarative memory that are missing from the Collins (2018) model, like forgetting or decay that happen in the span of hours, days, and years.

Perhaps even before Ebbinghaus' famous late 19th century memory savings studies, memory decay has been a critical component of how we understand learning and memory. Declarative memories allow for quickly learning flexible associations, but they succumb to decay. But frequent and spaced retrieval of these associations, before they become irretrievable, protects specific declarative memories from decay, and they become more accessible (Anderson, 2000).

Declarative memory decay is formally represented in computational models by the distinctive power-law function. In simple terms, a power law function, which defines the relationship between time and a measure of recall (to keep it in context, accuracy, for example) will appear linear if we take the logarithm of both variables (Newell & Rosenbloom, 1980; Anderson, 1982; Anderson, 2000). Therefore, a measure of memory decay should be a part of individual cognitive characteristics to examine for learning. In more recent years, model parameters in the ACT-R cognitive architecture (Anderson, 1982; Anderson, 2007, detailed in chapter 2) that represent memory decay have also captured individual differences (e.g., Van Rijn et al., 2009; Sense et al., 2021; Hake & Stocco, 2023), suggesting that we all forget information at different rates. In summary, along with working memory capacity, memory decay rate is used as a measure of declarative memory performance and its relation to strategy is investigated.

Lastly, on the other side of this dichotomous view, the procedural or non-declarative memory side, learning occurs by trial-and-error and repeated practice, as described in the earlier skill learning section. The basal ganglia, especially the striatum, are involved in procedural learning but there are multiple accounts as to how the striatum supports and shapes automatic stimulus-response maps with repeated practice (see Stocco et al., 2010 for a review). Several studies have shown that trial-and-error or learning by exploration, is reward-based, and is driven by dopamine signals from the striatum that are sensitive to the predictive utility of current actions and decision on future rewards, called reward prediction errors (e.g., Schultz,1998; Niv, 2009). In other words, behaviors and decisions have more value, and are reinforced, if they are more likely to lead to a reward in the future (Sutton & Barto, 1998) . And as described above, patients with Parkinson's Disease do poorly on procedural memory tasks like motor skills and probabilistic categorization. Trial-and-error learning is captured by reinforcement learning (RL) and a family of RL models (see Chapter 2, Equation 1).

There are few examples of individual differences of procedural memory but models of reinforcement learning as in Collins (2018) above rely on a learning-rate parameter (described in more detail in the following chapter) that could be allowed to vary freely to capture individual differences. Individual differences in basal ganglia function and its relation to behavior have been captured by imaging and computational studies. One such study found that RL model estimates of reward prediction errors, (a model that includes learning rate as a parameter) was significantly related to both behavioral outcomes on a probabilistic learning task and fMRI activation patterns in the striatum (Schonberg et al., 2007). In summary, differential patterns of striatal activity drive trial-and-error learning success, and these are correlated with model parameters that best fit learning behavior.

Brief Introduction to the Current Study

The current study is structured around three main questions that were designed to address the hypotheses laid out thus far. We adopted a simple stimulus response learning task called the Reinforcement Learning Working Memory Task (RLWM) from Collins (2018), to test our hypotheses. This task was chosen because it presented a learning environment that was designed to dissociate behavior guided by working memory (what we understand as declarative memory), measured by an ‘easy’ condition, from procedural (RL) memory, measured by a ‘difficult’ condition . These two conditions simply place varying demands on working memory capacity.

First, given that individuals might be using different learning strategies can we capture and identify these strategies on the RLWM task? To answer this question, we fit multiple computational models that each represent a different strategy and see which model best describes the behavior of our participants. It is crucial here that we fit models to each individual separately, unlike other modeling efforts that fit group-level behavior, so we can determine individual differences. The strategy models, task, fitting procedure, and justification for this approach is detailed in the next chapter (Chapter 2).

Second, if we are able to identify strategies we seek to further understand if they are stable in a learner across time, and in response to task demands. This is important since, as outlined above, different tasks demands might influence learners’ approach. They also likely engage in metacognition (McDonough et al., 2021) which might lead to changes in strategy. Here, we fit each participants’ behavioral data at two different time epochs to see if the same models explain behavior. Additionally, as described above, we test if different models explain participants' behavior on two conditions that represent the two levels of difficulty described above.

In order to draw a level of confidence that our methods identify true strategies we performed a model and parameter recovery analysis. Simply, this involves fitting simulated participants with known strategies and parameter values and checking how much of that ground truth information is correctly identified by our models.

Third, we sought to find a connection between strategies, learning outcomes and parameter estimates. In chapter 4, we rely on model parameters for declarative memory and reinforcement learning to capture latent cognitive characteristics on the RLWM task. But we went a step further to test if these parameter estimates are more general cognitive measures. If so, we expected that these parameter estimates would be stable across tasks. To that end, we included two new tasks and their associated ACT-R models to test if these parameters were correlated across tasks. The details of the tasks used along with explanations of parameters are presented in Chapter 4.

This report is organized around the three questions above and presented in separate self-contained papers, two of which have been presented and published at conference proceedings.

Chapter 2: One Size Doesn't Fit All: Idiographic Computational Models Reveal Individual Differences in Learning and Meta-Learning Strategies

Theodros M. Haile

Chantel S. Prat

Andrea Stocco

Department of Psychology, University of Washington Seattle

Abstract

Introduction. Complex skill learning depends on the joint contribution of multiple interacting systems: working memory (WM), declarative long-term memory (LTM) and reinforcement learning (RL). The present study aims to understand individual differences in the relative contributions of these systems during learning. **Methods.** We built four idiographic, ACT-R models of performance on the Collins (2018) stimulus-response learning, RLWM task. The task consisted of short 3-image, and long 6-image, feedback-based learning blocks. A no-feedback test phase was administered after learning, with an interfering task inserted between learning and test. Our four models included two single-mechanism RL and LTM models, and two integrated RL-LTM models: (a) RL-based meta-learning, which selects RL or LTM to learn based on recent success, and (b) a parameterized RL-LTM selection model at fixed proportions independent of learning success. **Results.** Each model was the best fit for some proportion of our learners (LTM: 68.7%, RL: 4.8%, Meta-RL: 13.25%, bias-RL:13.25% of participants), suggesting fundamental differences in the way individuals deploy basic learning mechanisms, even for a simple stimulus-response task. Finally, long-term declarative memory seems to be the preferred learning strategy for this task regardless of block length (3- vs 6-image blocks), as determined by the large number of subjects whose learning characteristics were best captured by the LTM only model, and a preference for LTM over RL in both of our integrated-models, owing to the strength of our idiographic approach.

Keywords: Individual differences; learning; ACT-R; reinforcement learning; working memory; declarative memory

Introduction

Individual differences in the ability to learn new associations are foundational to most measures of aptitude— a construct that describes the readiness with which one can acquire a complex skill. But even basic associative learning paradigms, like stimulus-response mappings, have been shown to rely on a mixture of cognitive mechanisms including working memory, reinforcement learning, and declarative memory (e.g., Poldrack et al., 2001; Stocco et al., 2010). Though a considerable amount of research has investigated how task characteristics drive the deployment of these mechanisms during learning (e.g., Collins & Frank, 2012; Poldrack and Packard, 2003), less work has been devoted to understanding how and when they may be differentially deployed across different individuals. To examine this, we built four models of a stimulus-response learning task (Collins, 2018) using the Adaptive Control of Thought - Rational (ACT-R) cognitive architecture (Anderson, 2007), which relied upon different combinations of learning mechanisms, with the goal of characterizing the specific learning strategies deployed by individual learners.

Multiple, parallel memory systems, notably procedural and declarative memory systems, facilitate how a subject learns and responds to their environment (e.g., Cohen et al., 1998; Ullman, 2001; Squire, 2004). The specific requirements of the task often dictate which memory system acquires information and guides behavior (McDonald and Hong, 2013). Some tasks require slow learning of response patterns through trial and error and repetition with feedback. This results in specialized, efficient skills that are learned typically through procedural memory (implemented through reinforcement learning, Niv, 2009). Examples extend beyond most motor skills, to procedural bases for cognitive skills like language (Ullman, 2001) and mathematical problem solving (e.g., Anderson, 1982). Performance of such procedural skills are usually not susceptible to distractions, suggesting relative disengagement from cognitive resources like working memory. They are, however, difficult to interrupt and modify during execution (e.g., Hikosaka et al., 2013), if, for instance, the appropriate responses in the environment change (McDonald and Hong, 2013).

Other tasks might benefit from the ability to learn new associations, facts, and categories rapidly, a process ascribed to the declarative memory system. This system is closely tied to, and affected by, attention and working memory (e.g., Engle, 2002; DeCaro et al., 2008). These types of skills are susceptible to distraction and are strained by high-load, difficult tasks. Some examples are learning to perform a new task via instructions or forming arbitrary associations.

Multiple lines of evidence show that these two memory systems often interact, and even compete, for task control (e.g., Collins, 2018; Poldrack et al., 2001); and individual differences research suggests that which of the two systems 'takes charge' of behavior is not determined by the task alone (e.g., McDonald and Hong, 2013;). An individual's cognitive capabilities such as working memory function (Just and Carpenter, 1992; DeCaro et al., 2008), declarative memory (e.g., Gluck, 2002), and procedural skill (Kalra et al., 2019) learning may also shape the extent to which one, or the other, memory system is deployed during learning. These individual characteristics are sometimes stable and result in the same learning outcomes across different sessions of the same task (e.g., Kalra et al., 2019, on both procedural and declarative tasks) but may also vary across tasks (Knowlton & Squire, 1996). These differences are especially apparent in complex tasks that seemingly draw on both declarative and procedural mechanisms, like the popular Weather Prediction Task (Gluck, 2002).

However, it is not well understood when and why these differences arise. They may arise due to any combination of meta-cognitive awareness, learning experience, level of expertise or previous knowledge, and individual cognitive capabilities like working memory capacity or forgetting rate. For instance, learners with high working memory capacity learned categories more slowly, in a task where item categorization was based on multiple relevant features, and procedural learning systems were better suited for the task (DeCaro et al., 2008). This suggests that these high working memory capacity learners tried to learn the categories declaratively by maintaining object features in working memory, and that this strategy was suboptimal. A critical question that arises is, how do individual learning approaches interact with task requirements to produce successful outcomes? To address this question, we adopt an individual differences approach to the study of associative learning.

Many studies examine memory systems, and their neural correlates (e.g., Squire, 2004; Poldrack et al., 2003; Puig et al., 2014), but very few study how or why different learners might arbitrate between and deploy different memory mechanisms for a given task. In this one-size-fits-all approach, very few studies consider individual differences. In fact, most experiments are designed to reduce between-subject variability and maximize group effects, which makes many popular learning tasks unsuitable for individual differences studies (Hedge, 2018).

In the current study, we sought to address some of these differences in learning by fitting and comparing multiple, single-system, and multi-component models to individual participants' data. The two single-system models were designed to capture sole reliance on one learning system alone, while our multi-component models were designed to capture two ways of integrating them: flexibility in mechanism selection that arises from meta-cognitive sensitivity to their most recent performance, or a biased but stable preference for a mixture of learning mechanisms, suggesting cognitive entrenchment resulting from the learners' history and experience with learning. These models were built in the ACT-R cognitive architecture (Anderson, 2007) discussed in the following section.

To address the questions outlined above, we chose the Reinforcement Learning Working Memory (RLWM) task (Collins, 2018), because it has previously been used to dissociate the contributions of different learning mechanisms. While we rely on this versatile task for our modeling efforts, our assumptions about learning systems, their interactions and model fitting procedures depart from Collins' (2018) in several ways, discussed subsequently.

In the RLWM task, participants are asked to learn associations between images (e.g., objects, shapes, and colors) and one of three potential keyboard presses using feedback. Collins originally developed the task to quantify the relative contributions of working memory and reinforcement learning through two training conditions. The first condition consisted of short blocks where three stimulus-response associations were learned, and the second consisted of longer blocks where six stimulus-response associations were learned. Collins (2018) posited that the short, 3-image blocks would likely be learned faster and more efficiently through maintenance in working memory. The long blocks, on the other hand, would overwhelm working memory capacity limitations, making the system unreliable. Therefore, Collins (2018) proposed that they would rely more on reinforcement learning mechanisms, which are not capacity limited. To evaluate the extent to which working memory or reinforcement learning mechanisms were used, the task also included a surprise post-test after a 10-minute distracting task (the n-back task was used; see Figure 1). If working memory was primarily used to guide responses, the learned associations would decay during the intervening 10-minute distracting task and produce low accuracy during the test. If, on the other hand, the stimulus-response associations were learned primarily through reinforcement learning mechanisms, they should survive the break and produce high accuracy for the post-test. This assumption is supported by previous work examining the durability of reinforcement learning (e.g., Niv 2009; Stocco et al., 2010).

Computational modeling provides a robust approach for isolating the relative contributions of different learning mechanisms — which may be difficult to isolate with behavioral data alone (e.g., Stocco et al., 2021a; Stocco et al., 2021b; Collins, 2018; Stocco et al., 2017; Daw, 2011). Collins (2018) demonstrated how reinforcement learning (RL) and working memory (WM) may interact to learn the object-letter associations using a combined WM - RL model (RL+WMinteracting). They hypothesized that the WM resource, which is limited in capacity and decays rapidly but has a high learning rate, cooperatively interacts with the RL portion of the model, directly influencing the computation of the reward prediction error. When the number of images is high, as in the long blocks described above, the WM component of the model contributes less to reward prediction error. This interacting RL+WMi model fit participants' data best compared to other, RL only and non-interacting RL+WM models.

However, Collins (2018) did not consider contributions of long-term declarative memory processes, which may also have created durable representations of stimulus-response associations. Collins's (2018) original modeling effort implicitly assumes that all long-term associations between stimuli and responses are stored in a procedural, RL-based system, and, conversely, that all the explicit representations of the correct responses must fit within a temporally constrained working store. This is apparent in the assumption, for example, that performance after a 10-minute interval must reflect the RL system only (Collins, 2018).

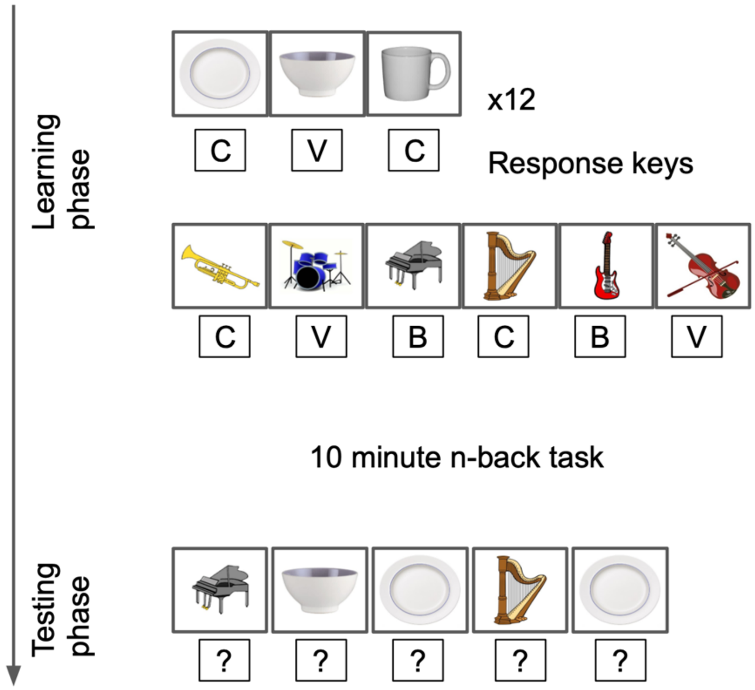


Figure 1: Schematic of the RLWM task. The images are examples of actual stimuli used in the task. A 1-back task was used as the distractor task during the 10-minute break before the testing phase. In this version of the task, two images were always associated with a letter in the set-size 6 block, and 2 images always had the same response in the set-size 3 block.

Additionally, many contemporary theories think of working memory as a process that arises from the interaction between attention and the strategic retrieval of long-term memory information (Kane et al., 2001; Miller, Lundqvist, & Bastos, 2018). Collin's (2018) modeling efforts confound the temporal axis of learning (long- vs. short-term representations) with the learning mechanism (implicit and procedural, driven by RL, and explicit, driven by WM).

Lastly, and critically for the current effort, Collins (2018) does not take account of individual differences in the deployment of learning mechanisms. They use a group-level model-fitting procedure that is tolerant of individual differences but did not try to fit different models to individual participants. Individual differences in both WM (Engle, 2002) and RL (Frank et al., 2007) are well documented, and these differences impact learning outcomes, especially when the to-be-learned tasks tend to be more complex.

In summary, while Collins (2018) provides evidence for the interaction between multiple learning mechanisms that aligns with other behavioral and neural evidence (e.g., Anderson, 1982; Poldrack, 2001; Antzoulatos and Miller, 2014), the study de-emphasizes the individual differences that might dictate how these systems interact. We argue that this, and the failure to consider longer-term declarative learning strategies (Poldrack, 2001; Schneider and Chein, 2003), creates an incomplete account of the cognitive bases of associative learning. Therefore, we have made declarative memory and individual model fitting procedure central to our experimental efforts.

Modeling individual differences in learning using ACT-R

To capture the interplay between RL, declarative memory (LTM), and WM in individual learners, we built a series of models using the ACT-R cognitive architecture (Anderson, 2007). ACT-R was an obvious choice for this study because of its expansive, flexible, and manipulable integration of learning mechanisms. In ACT-R, knowledge is represented in two possible formats: procedural and declarative. Procedural knowledge is represented as procedural rules whose utility is learned through RL (Ceballos et al., 2020; Stocco et al., 2010). Declarative knowledge is represented as explicit memories in a structured record format. Explicit memories decay over time following the power laws of recency and frequency (Anderson, 2007). Their activation, however, can be momentarily increased through spreading activation, an attentional mechanism that can be used to maintain information for a brief amount of time and predicts individual differences in working memory capacity (Daily et al 2001). Finally, ACT-R is a realistic “end-to-end” modeling tool that includes multiple modules to capture sensorimotor interactions with a task.

In the current study, we built four models to capture typical learning trajectories and outcomes in a declarative memory only system (LTM model) with a variable WM analog, an RL only system (RL model) and two combined RL, WM, and LTM models (RL-LTM models), discussed below.

Method

Participants

83 undergraduate students from the University of Washington participated in this experiment. All participants were monolingual English speakers recruited through the UW Psychology subject pool and on-campus posted advertising (47 females, aged 18-35 years). Data were collected after receiving informed consent in one 2-hour session.

Materials

Behavioral Task

The Reinforcement Learning Working Memory task (Collins, 2018) involves learning stimulus-response associations through a series of 14 blocks. Participants are instructed to respond with a keypress of either 'C', 'V' or 'B' to the displayed images. In 8 of the 14 blocks, participants learn to associate keypresses with three unique images, presented 12 times in random order. In the remaining 6 of the 14 blocks, they have to learn to associate 6 unique images each presented 12 times within the block, with those same letters. The stimulus-response associations are arbitrary but deterministic, and participants learn through trial-and-error and with explicit feedback (+1 point for correct responses and 0 points for incorrect responses). Images are shown for 2 seconds, within which time participants must provide a response, that trial is counted as a miss otherwise. The next trial begins automatically right after feedback. Following this learning phase, a 10-minute distractor, 1-back working memory task , is administered before a surprise 206-trial test block. Participants make responses without feedback to items taken from both the 3- and 6-set learning blocks. In the current version of the task, each response in the set-size 6 condition was always related to two images, and two images were always associated with the same response, as shown in Figure 1. Transitions between blocks were initiated by participants by keypress. Stimulus presentations and data collection were done in MATLAB (mathworks.com) and Psychophysical toolbox (psychtoolbox.org).

Computational Models

All the models experienced the same experimental set-up — 2 learning blocks of 3 and 6 objects, a 10-minute delay, and a test phase without feedback. As each block contained unique stimuli that were entirely new to the model, multiple blocks were not needed. Furthermore, 100 simulations were run for each parameter combination of each model.

Reinforcement Learning Model. The first model (Figure 2) most closely adheres to Collins' RL model. This model uses production rules to represent all of the possible stimulus-response associations and uses reinforcement learning to progressively learn which associations are correct. Each production rule p has an associated utility value, $U(p)$, that reflects its expected rewards and is learned through a temporal difference rule. Specifically,

$$U_t(p) = U_{t-1}(p) + \alpha\,[R_t - U_{t-1}(p)] \tag{1}$$

in which α is the learning rate and R_t is the reward given at time t. In our experiment, R_t is binary and corresponds to the feedback ("Correct", $R_t = 1$, and "Incorrect", $R_t = -1$) given by the task interface. Competing responses are selected on the basis of their respective utilities, using a soft-max rule controlled by a noise parameter τ. The model initially responds randomly, until the correct rule accrues sufficient rewards to overcome the competitors, given the noise τ. The entire RL model is controlled by those two parameters, the learning rate α and the selection noise τ.

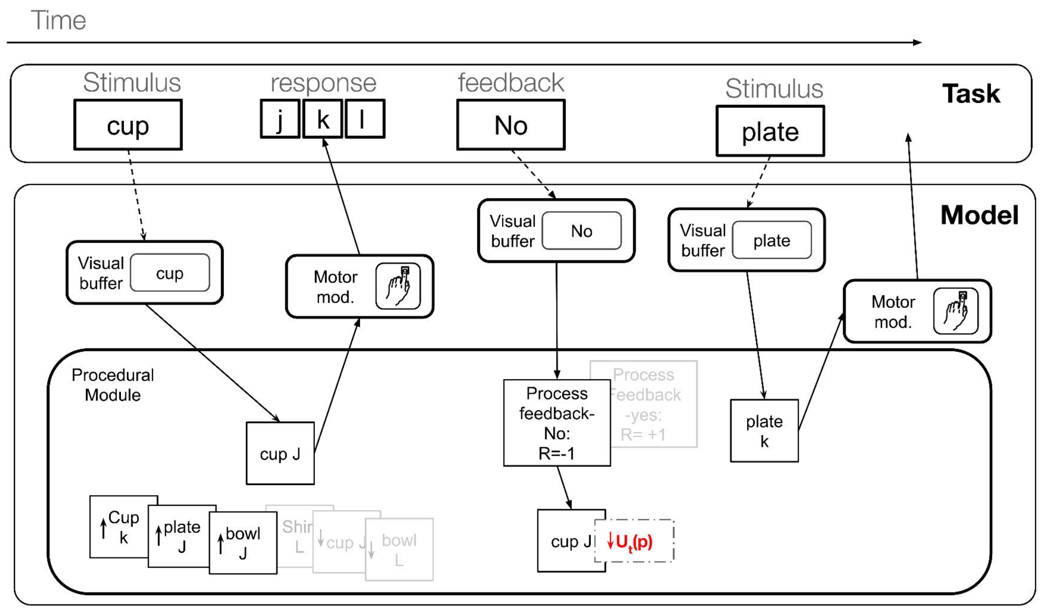


Figure 2: Overview of the procedural RL model as implemented in ACT-R. In this example case, when a picture of a Cup is presented, the model responds "J" by selecting the "Cup J" production. The feedback, 'No', is sent to the RL-based utility learning system in ACT-R which reduces the *utility* of the "Cup J" production. Changes in utility through learning are signified by up and down arrows in this diagram. Utility values for productions are updated only when they are selected for a response.

Declarative Learning Model. In lieu of Collins' pure WM model, we developed a declarative model (Figure 3) which manages both long-term and short-term explicit associations between a stimulus and its correct response. This model stores memories of specific task events for later recall and use. To start, the model attempts to retrieve a memory of a previous correct response to the current stimulus. If such a memory is found, the same response is used. If no memory can be found, the model makes a random response. The response to the current stimulus and its outcome (correct or incorrect feedback) are then memorized. Although this model is computationally simple, ACT-R allows for a sophisticated control of the memory management processes through three parameters (Table 1): (a) activation noise *s*, which captures random fluctuations in a memory's activations and associated probability of retrieval, (b) decay rate *d*, which captures the rate at which memories fade away and are forgotten (Sense et al., 2016); and

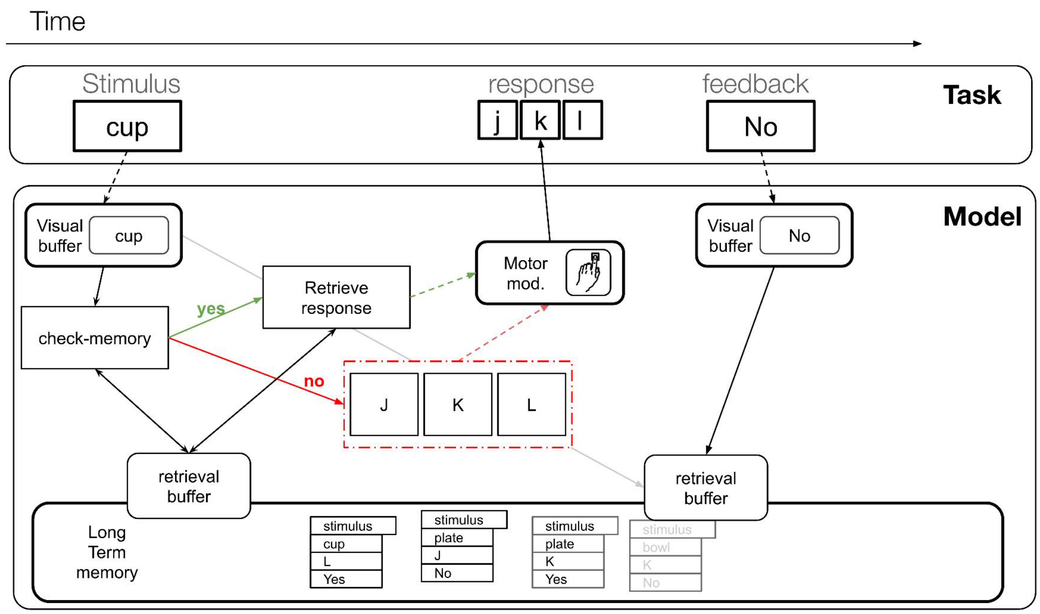


Figure 3: Overview of the declarative model as implemented in ACT-R.

(c) spreading activation weight *W*, which captures the attentional resources allocated to activating relevant memories during retrieval and has been shown to capture individual differences in working memory capacity (Lovett, et al., 2000; Daily et al, 2001). We hypothesize that individual differences may occur in this three-parameter space and might be an intrinsic source of strategy choice during learning and retrieval.

Integrated LTM-RL models. Our third and fourth models (Figure 4) integrate the two single-system models into two new multi-component RL – LTM models with differences in trial-by-trial arbitration and selection of a sub-system for engagement. Both models initiate each new trial by first deciding which of the two strategies to use — the procedural or the declarative strategy. The mechanism for integration provided a specific challenge. What is the

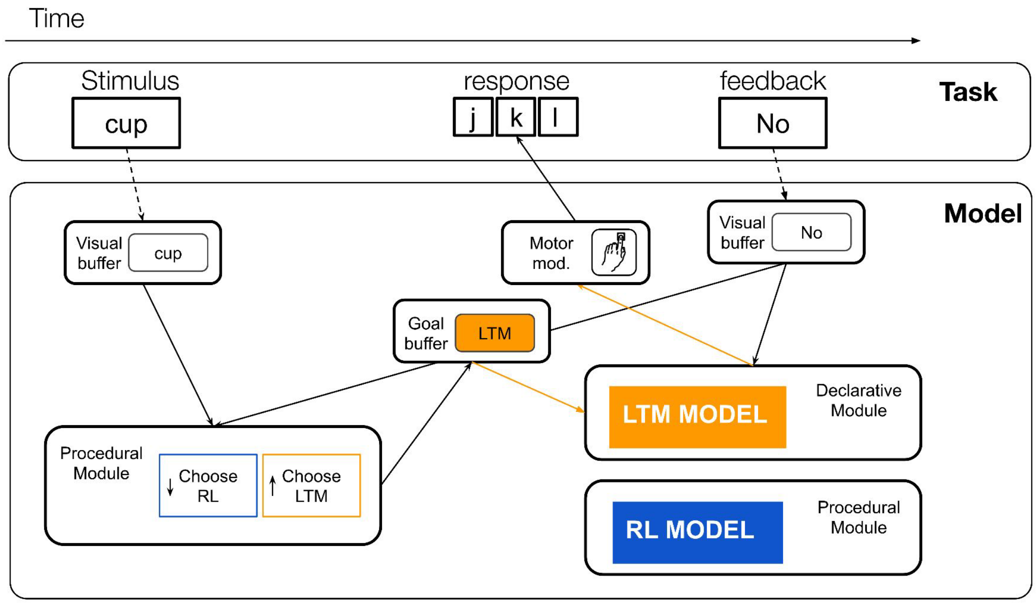


Figure 4:An overview of the integrated Meta-RL model. The meta-learning model implements a utility learning component that selects the most successful sub-model (RL vs LTM). The figure above exemplifies a scenario where the LTM sub-model was selected. Up and Down arrows in productions signifies continued utility learning.

most likely way that these two systems collaborate or compete during learning and recall? We decided to test two possible ways a meta-learner could arbitrate which system to use. The first (Figure 4), perhaps more elegant, solution was to have a reinforcement learner that learned the best strategy given the specific set of parameters (Meta-RL model). This model has five parameters total, the two inherited from the pure RL model (α and τ) and the three inherited from the Declarative model (s, d, and W). This model assumes that individuals are adaptive learners and can optimally choose strategies based on their relative success over a short time. For example, if the long-term memory strategy proves too difficult (as in the case of too many stimuli), the model would switch to a RL-based learning strategy. RL learned associations are shared with the LTM system by inserting explicit information into the memory module. The meta-learner's proportions of RL vs LTM selection is determined at the end of each simulation, for each parameter combination. This allows us to measure what the combined effect of parameter values was on RL vs LTM preference at the end of learning.

The second integrated model (Biased Model, Figure 5), has a built-in preference bias towards one system, quantified as a bias parameter β. Thus, at the beginning of every trial, the model selects the procedural/RL strategy with probability β and the declarative strategy with probability 1 - β. In contrast to the previous model, this bias is fixed and does not change over the course of the task. The Biased model embeds the hypothesis that individuals might have established preferences towards one way to learn or another, perhaps honed over many years of "learning to learn" across contexts and circumstances. For instance, if an individual prefers declarative learning, they will persist in trying to memorize stimulus-response associations even when switching to a RL strategy would be more convenient.

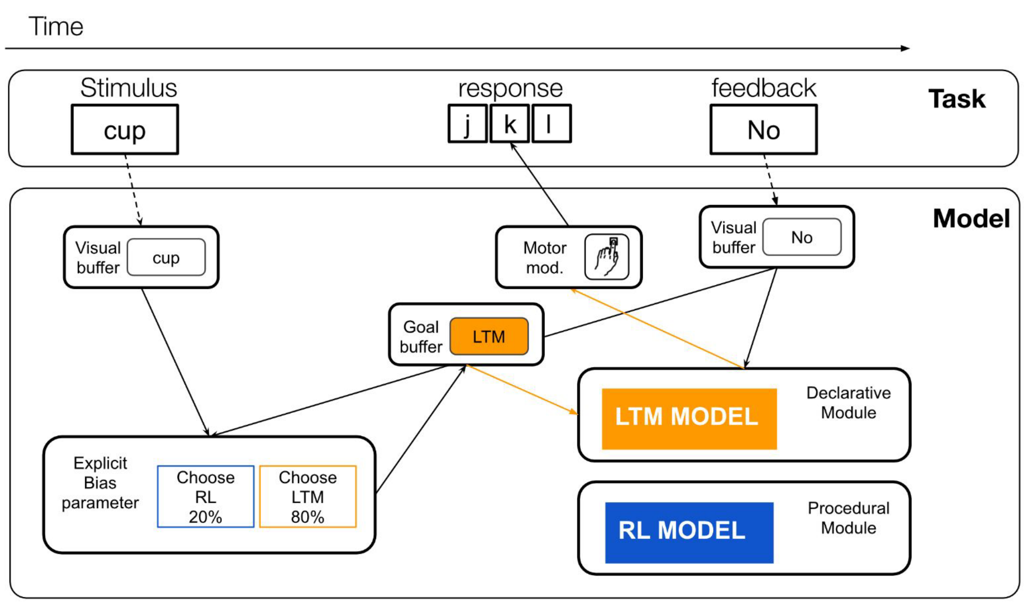


Figure 5: An overview of the integrated Biased model. The explicit Biased model does not use the procedural module utility learning but instead selects either RL or LTM at predefined proportions of either 20%, 40%, 60% and 80%. The specific example above shows a time-point where the bias parameter prefers RL only for 20% of the time.

Simulations

In this study, models are used as investigative tools to better characterize each individual. To do so, each model was run across a discretized version of its parameter space. Despite being computationally expensive and coarse, this method was preferred to convex optimization methods because it gives the full view of parameter space (including local and global minima) and, once computed, does not need to be recalculated for each participant (Fisher, et al., 2016). To obtain stable estimates, each model was run 100 times for each possible combination of parameters. In discretizing the range of each parameter, values were chosen to form an interval that surrounds the recommended value in the ACT-R documentation. The spreading activation parameter values however were selected further away from the recommended value of 1 because a value of 1 and above injected more than sufficient spreading activation with no room for effect variability. A full description of parameters and the range of values that were manipulated is given in Table 1.

Table1

Parameters and Range Explored for Procedural (RL), Declarative modules (LTM) and integrated (RL-LTM) Biased Models.

	Parameter	Values				
RL	alpha(α)	0.05	0.1	0.15	0.2	0.25
	softMax (τ)	0.1	0.2	0.3	0.4	0.5
LTM	decay rate (d)	0.3	0.4	0.5	0.6	0.7
	activation noise (s)	0.1	0.2	0.3	0.4	0.5
	spreading activation (w)	0.1	0.2	0.3	0.4	0.5
RL-LTM	bias (β)	0.2	0.4	0.6	0.8	

Data Analysis and Participant Fitting

Each participant's meta-learning strategy and latent, idiographic characteristics were then measured by identifying the model that best reproduced their observable data Y. Specifically, each participant was matched to a particular model *M* and set of parameter values θ_M, that minimized the following function:

$$M, \theta = \text{argmin BIC } (Y_p, Y_M \mid M, \theta) \tag{2}$$

in which Y_p is the observable task performance from participant *p*, Y_M is the simulated task performance, *M* is one of our four given models, θ_M is its associated set of parameters, and BIC is the Bayesian Information Criterion (Schwarz, 1978), which can be further expressed as:

$$\text{BIC} = n + n \log (2\pi) + n \log (RSS)/ n) + \log (n) (k + 1) \tag{3}$$

in which *n* is the number of data points to fit, *k* is the number of parameters in each model, and *RSS* is the residual sums of squares. In our case, the n data points are the 24 mean accuracies associated with the presentations of each individual stimulus (12 for set-size 3 and 12 for set-size 6), plus the two post-learning test accuracies.

The BIC was chosen because it incorporates both fit and model complexity in a Bayesian framework, thus natively accounting for the fact that a more complex model has an a priori greater likelihood to fit a given individual and that, given two models that fit the same data equally well, the one with the smallest number of parameters is the more likely to the be best model for that individual.

Results

Overview of Behavioral Results

By and large, our behavioral experimental results replicated the experimental findings of Collins (2018). On average, participants' performance improved throughout the learning phase of the experiment (Figure 6A), as shown by a significant effect of the stimulus repetition on its response accuracy ($F(11,1968) = 412.911$, $p < 0.001$). As previously reported, stimuli in the set-size 3 condition were generally learned faster (learning rate: $t(142.6) = 10.15$, $p < 0.001$) and better than those in the set-size 6 condition (accuracy at end of learning: ($W = 4025.5$, $p = 0.057$; Figure 6A). Finally, the two conditions (set-size x learning and test phase) interacted ($F(1,328) = 8.14$, $p = 0.0046$; Figure 6B), with greater forgetting in set-size 3 as compared to the set-size 6 between training and test (Figure 6B).

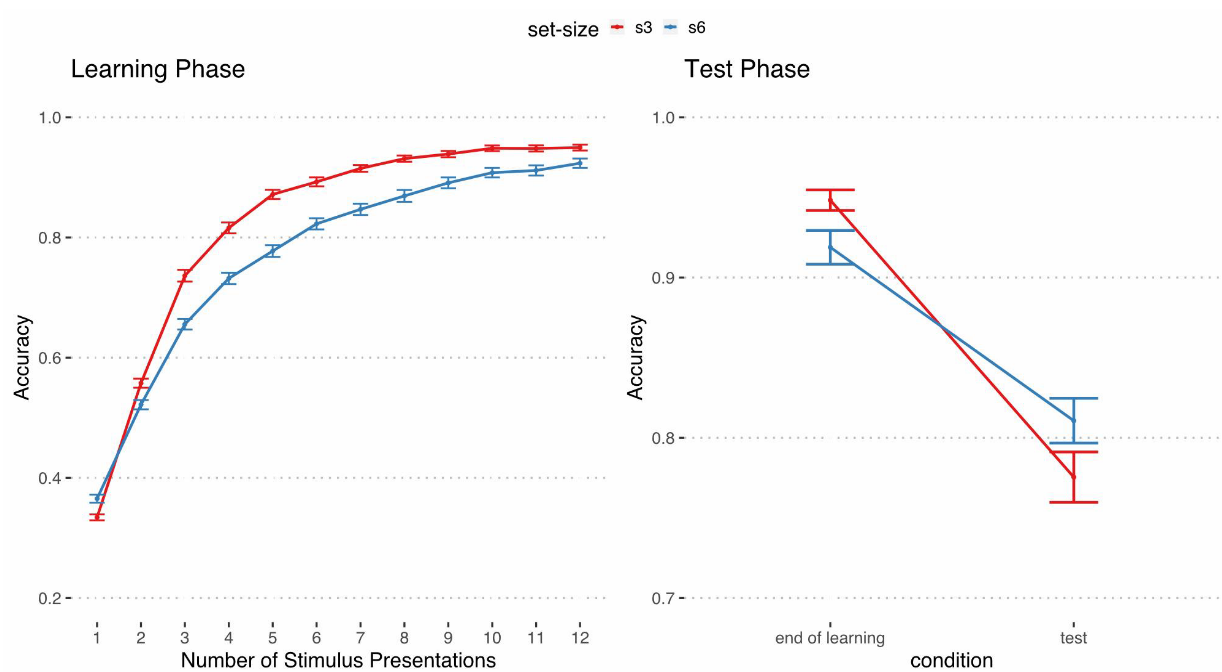


Figure 6: **(A)** Accuracy across successive stimulus presentations during the RLWM task and the **(B)** Change in accuracy from asymptotic learning (last 5 stimulus presentations) to the test phase in the RLWM task.

These group-level results suggest that individuals use a mixture of declarative and procedural strategies. This is shown by the different effects of the 10-minute distracting break on the two set-sizes during the testing phase. It appears that some information has decayed over time for set-size 3 objects, possibly compatible with declarative memory and working memory utilization. Information was preserved for set-size 6 objects, which aligns more with a procedural memory utilization, but also, possibly, a declarative memory strategy. Additionally, the superiority in speed and accuracy of stimulus-response learning of objects in the set-size 3 condition rules out reinforcement learning.

Overview of Model Simulations

The four models displayed different learning trends for the same tasks, even when mean performance across the entire range of parameter space was taken (Figure 7), which suggests that different learning strategies alone produce variable outcomes. As Collins (2018) pointed out, the pure RL model predicted no difference between set-size 3 and set-size 6. Notably, our pure LTM model also predicted a very minimal difference between the two set-sizes, at least within the range of our tested set of LTM parameters. The mixture models, however, predicted differences between the two set-sizes, with the difference being stronger for the explicit, Biased model. Analysis of the Meta-RL model suggests that this might be a side effect of the model using different strategies for set-size 3 and set-size 6 stimuli, due to an interaction between set-size and specific parameter values. Recall that the parameter values influence the success rate of the sub-models (see *Capturing Individual Differences* below for details). Additionally, the LTM model had the fastest learning rate of the four models followed by the Meta-RL model, then the pure RL model, and the Biased model last.

The models also differed in accuracy at the end of learning. Here, again, the LTM model presented the most success by attaining close to 100% accuracy at the end of learning irrespective of set-size. The rest of the models followed the same trend as the learning rate with the Meta-RL model achieving 87% (set-size 3) and 90% (set-size 6) accuracy, the pure RL model around 77% for both set-sizes, and the Biased model at 79% (set-size 3) and 75% (set-size 6) accuracy.

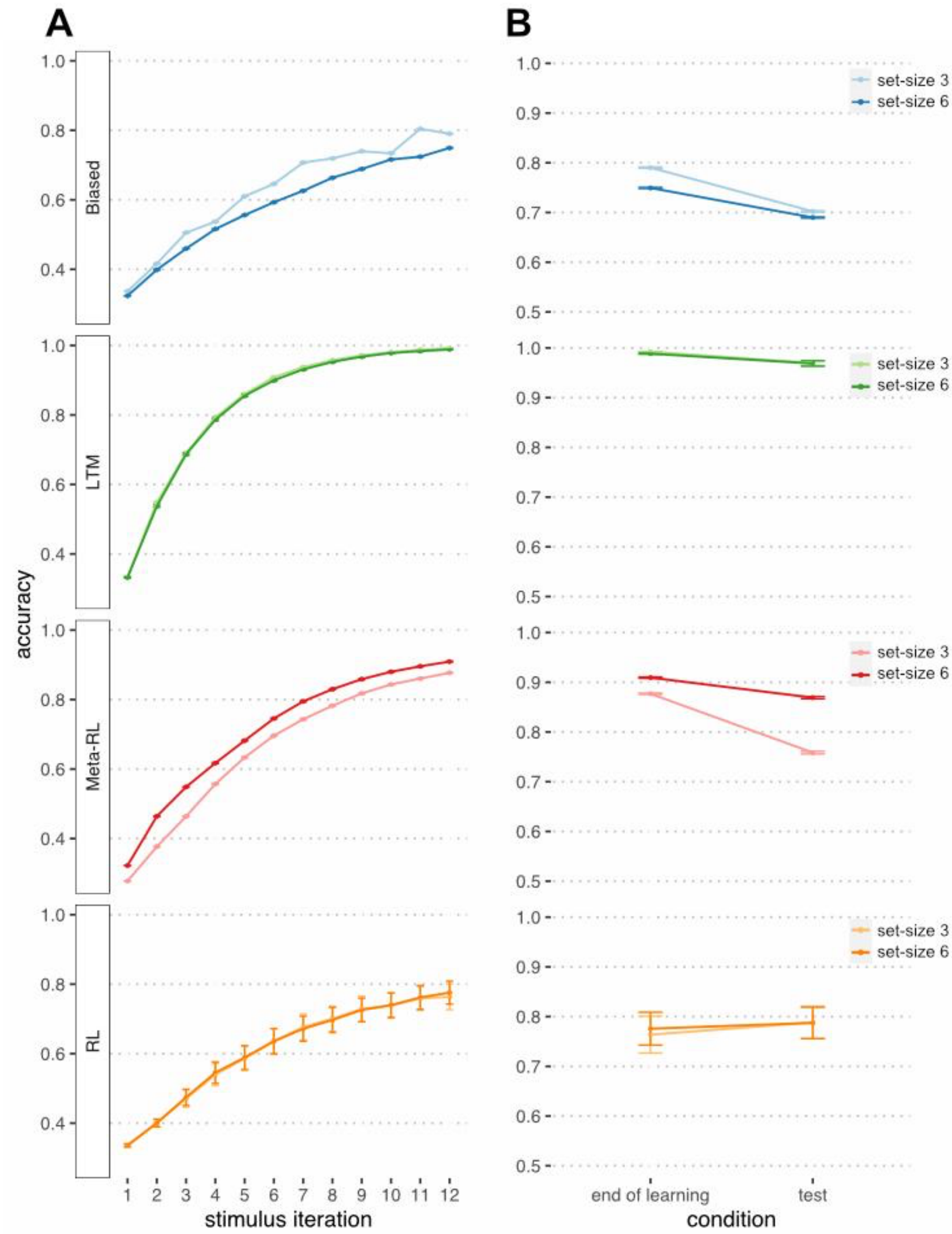


Figure 7: Model Learning accuracies for successive presentations of stimuli (A) and accuracy at end of learning, and test (B). The data points are mean accuracy averaged across all parameter-sets. Light colored lines, points and curves represent set-size 3 and dark colored points, lines and curves represent set-size 6. Each row shows simulation data from our four models: 'Biased' figure shows the learning and test for the explicit bias strategy integrated model (12,500 parameter-sets), 'LTM' figure shows learning and test for the pure LTM model (125 parameter-sets), 'Meta-RL' figure shows learning and test for the meta-learning RL integrated model (3125 parameter-sets), and the final figure at the bottom shows learning for the pure RL model (25 parameter-sets).

Individual Model Fitting Outcomes

After examining the behavioral results, each participant was matched to an ideal model using the BIC minimization procedure described above. To assess the reliability and stability of the model fitting procedure, the BIC values for each participant-to-model fit were ranked (Figure 8) and compared by taking the differences in BIC for each consecutive best-fit model. A difference of 6 to 10 suggests strong evidence for the model with the lowest BIC value; differences between 2 and 6 suggest positive evidence, and a difference less than 2 suggests weak evidence (Raferty, 1995).

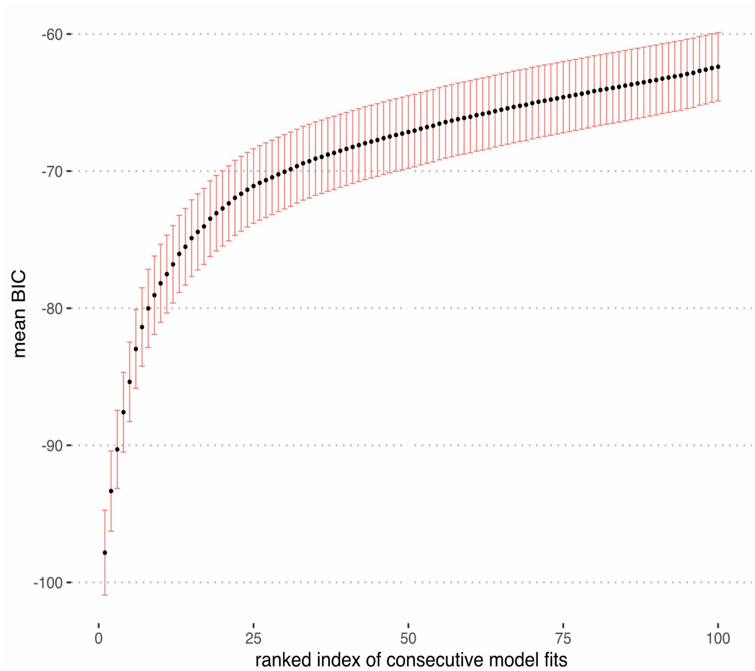


Figure 8: Mean BIC values for ranked models fits, therefore, the first data point shows the mean BIC of the best fit model, which might be different for most participants. Subsequent models could be from the same model family, with different parameter values or entirely different models. Data points show mean BIC and error bars show standard errors across participants.

We found that the difference in BIC value for the best fitting (M_1) and the second-best fitting model (M_2) was in the positive evidence range ($\Delta M_{1,2}$ = 4.49; *SEM* = 0.486). This difference fell when comparing the 2nd with the 3rd best fitting models ($\Delta M_{2,3}$ = 3.04; *SEM* = 0.355) and 3rd with 4th ($\Delta M_{3,4}$ = 2.71, *SEM* = 0.323). These results indicate that the best fitting model selected for each participant has good evidence of fit, at an estimated 75 to 95% posterior probability, against the subsequent models given the data (i.e., $P(M_1|D)$, Kass and Raferty,1995). When split by model type, cases where the RL model fit participants best, demonstrated the strongest evidence against the 2nd best fit model compared to the other models. However, there were only four participants that fit this model, so this metric is less reliable. The next best model is the LTM only model. Since a large number of participants fit this model, it might have the highest overall likelihood of capturing learning behavior in the RLWM task, even while it is not the model with the least number of parameters.

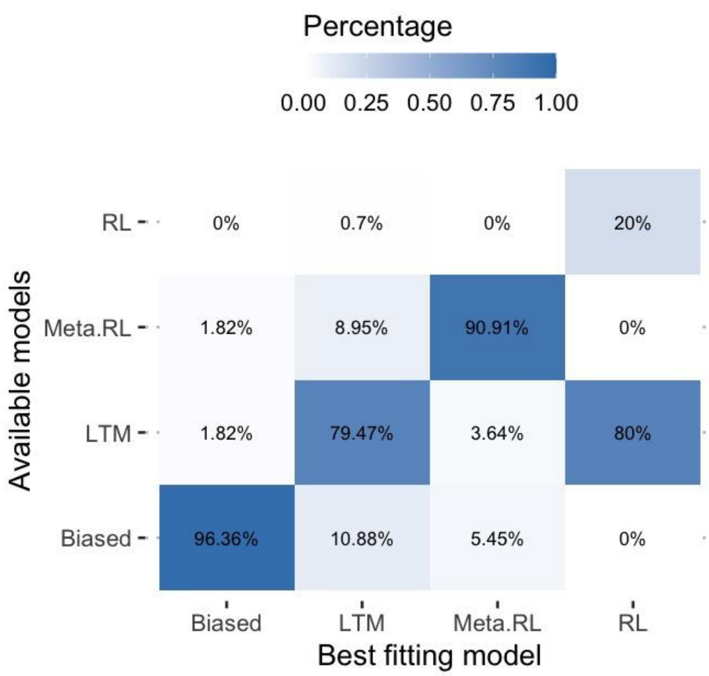


Figure 9: Confusion matrix showing proportions of fit models in the first, ranked, 10 best fit models.

Next, to test if the participants' data stably fit one of the 4 model types, and any differences were only due to changes in the parameters' values, we looked at how often a model from the same family was selected in the first 10, ranked, model fits(Figure 9). We found that the best-fitting model was selected on average 7.17 times out of 10. That is, if a model of one particular family was selected as the best-fitting model, 7.17 different parameterizations of the same model would also show up in the top ten. Specifically, the best-fitting model was selected 96.4% of the time for the Biased family of models, 79.5% for LTM, 90.2% for Meta-RL and 20% of the time for the four RL fitting participants. Seven participants (three from Meta-RL and all four from RL) had second best-fitting models that came from a different model family. This suggests that the results of the participant-model matching procedure reflect qualitative differences in the patterns of behavioral data rather than small differences in fit due to the discretization of parameters.

Capturing Individual differences

The goal of this study was to find a method for characterizing individual differences in the behavioral data. Therefore, collapsing across so much of this variability offered by the model design and set of parameters is uninformative. We proceeded with a 2-step approach to quantify variability in learning: first, we examined how each individual's best-fit strategy (model) explained their learning outcomes (discussed in this subsection) and, secondly, we analyzed the effects of parameters on learning outcomes and model performance. As discussed in the previous section, we found that different models steadily fit different subsets of participants, which makes the results of our model-fitting procedure somewhat different from the Collins (2018) study.

Firstly, of the four models compared, the LTM model fit most of the 83 participants (*n* = 57), followed by the Biased model and the Meta-RL model which tied for second place (*n* = 11). Only four participants best fit the pure RL model (Figure 10).

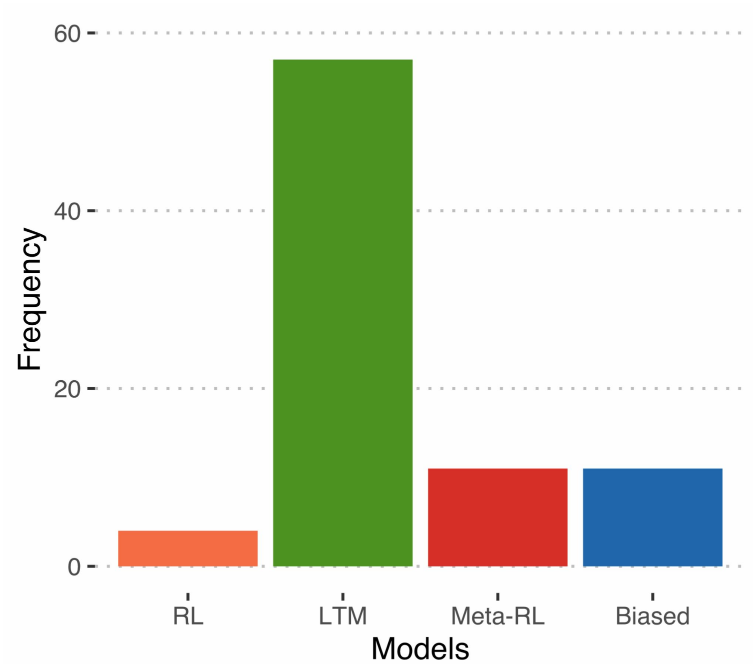


Figure 10: Counts of participants by best fit model group.

Secondly, we were interested in capturing the participants' specific behavioral outcomes, such as their learning rate, forgetting after break, and difficulty associated with the increase in the number of objects to learn (i.e., set-size differences). The models' performance matched some of these group-level behavioral performance in both the set-size and learning vs testing phase conditions.

We have shown that, on average, set-size 6 objects were remembered better than set-size 3 objects after the break (Figure 6B). Our models had captured this difference faithfully and revealed that the previously observed difference is true only for a subset of our participants, all of whom fit the Meta-RL model best (W =11.5, p = 0.0014; Figure 11B). It should be noted that the best-fitting Collins model (RLWMi) predicted higher forgetting in set-size 3 compared to set-size 6 during the test, after the break. However, the behavioral data from LTM group (W = 1386, p= 0.177), Biased integrated model (W =81, p = 0.19), and RL (W = 10.5, p = 0.53) show the same amount of forgetting during test for both set-sizes, which defies the WM – RL dichotomous view forwarded by Collins (2018). The Meta-RL group had also learned the images equally well at the end of learning, which was different from what Collins (2018) observed. This suggests that different strategies led to different learning outcomes, but also that our largest group likely used the same declarative strategy for both set sizes with robust effects that were not differentiable during the test.

Next, we observed that set-size 6 objects were associated with lower accuracy throughout the course of learning for most participants, suggesting increased difficulty, and this difference was significantly different from zero for the LTM ($t(56) = 6.8$, $p < 0.001$) and Biased ($t(10) = 10.37$, $p < 0.001$) group of participants, but not RL ($t(3) = 1.3$, $p = 0.28$) and Meta-RL ($t(10) = 0.90$, $p = 0.388$). Similarly, the learning curves from the best-fit model data showed significant differences between the two set-sizes, during learning for LTM ($p<0.001$) and Biased ($p<0.001$) groups but not RL and Meta-RL ($p = 0.08$).

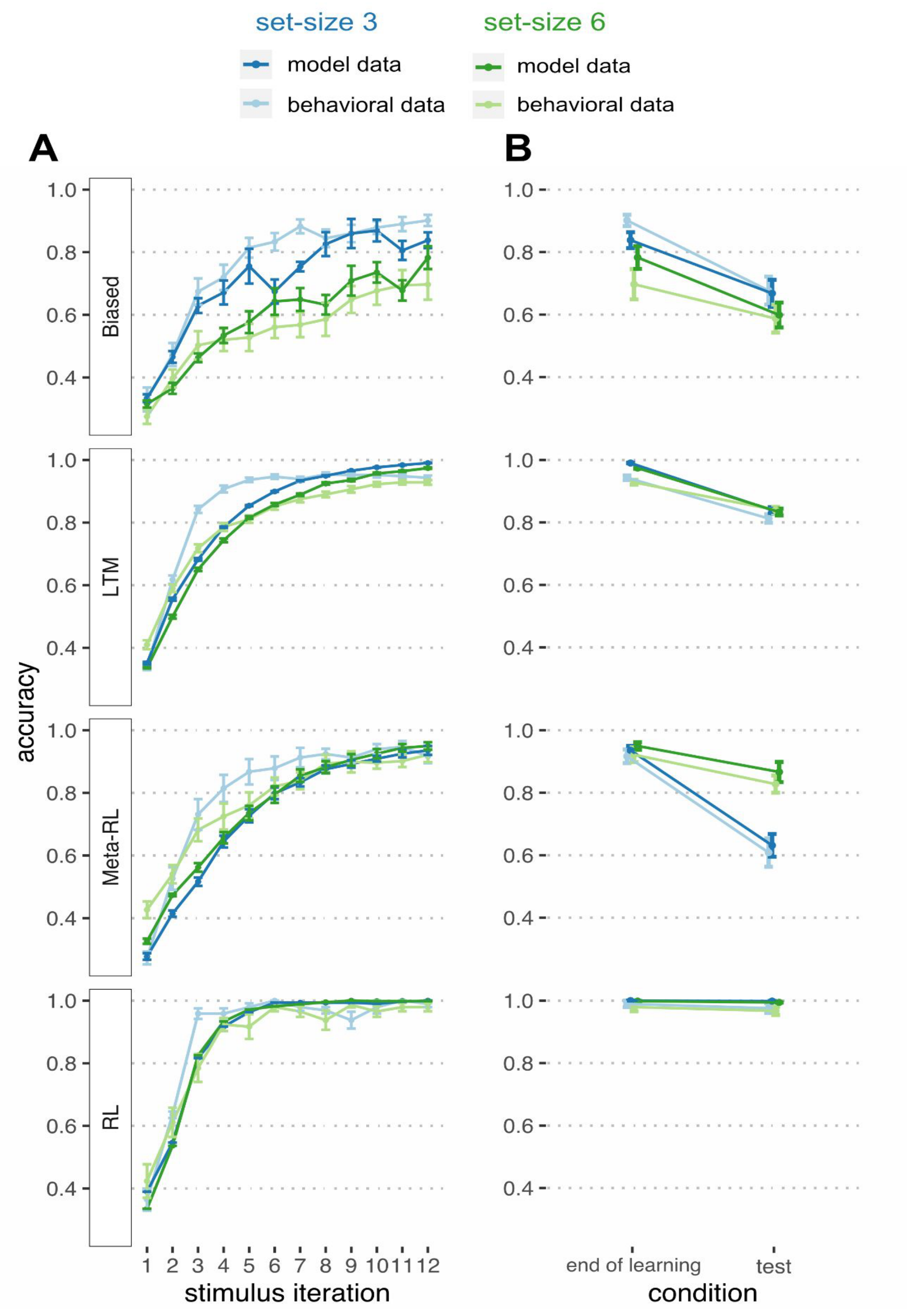


Figure 11: (**A**)Mean learning curves for set-size 3 (blue) and set-size 6 (green) objects.(**B**) Mean accuracy of learning in the last stimulus iteration, and testing. Light colors are averages

of subject data that fit that model best. Dark colors are averages of model data for parameter-sets that were best fits for the participants in that model group.

Examining learning rates (slope estimates from a linear fit to the first 6 stimulus iterations), for the four groups of participants, there was a significant main effect of group membership or best-fit strategy ($F(3,158) = 9.68$, $p < 0.001$), set-size ($F(1,158) =121.9$, $p <0.001$) but no interaction effects ($F(3,158) = 2.3$, $p = 0.07$). The four participants in the RL group had the highest learning rate compared to the other model groups for the set-size 3 and set-size 6 conditions (set-size 3: $M = 0.120$, $SEM = 0.0055$; set-size 6: $M = 0.110$, $SEM = 0.0084$) followed closely by the LTM and Meta-RL groups. These group differences were captured by the models, with better fit performance in our multi-component models (there were no main effect of data type, behavioral data vs model simulation, ($F(1,324) = 2.58$, $p = 0.10$); there was a significant main effect of model group ($F(3,324) = 29.5$, $p < 0.001$)), and there were no interaction effects between data and model type ($F(2,324) =1.4$, $p = 0.24$)).

Parameter analysis

Model learning was influenced by a maximum of 6 parameters ranging on five values (Table 1). The two integrated models, Meta-RL and Biased model utilized the five LTM and RL parameters, along with an additional bias parameter (β). In the Biased model, the bias parameter defines the proportion of RL vs LTM to use throughout the learning and test phases. For the Meta-RL model, we estimate an outcome variable that is similar to bias (β) by calculating the proportion of RL used by the meta-learner, for each set-size, at the end of the learning phase.

A select range of parameter values were explored to capture as much individual variability as possible, within the computational constraints of running simulations of all possible parameter combinations. After model fitting, a total of 36 parameter-value sets across all four models, out of 15,865 possible sets, described all 83 participants. Participants who fit the most popular model, LTM, that fit 57 participants, surprisingly, were described only by 14 parameter-value sets for the three LTM parameters out of the possible 125 combinations (spreading activation(*w*), retrieval noise(*s*) and memory decay rate (*d*)). The Biased model was the most diverse at 11 parameter sets for the 11 participants in that group (out of a possible 3125 combinations). The Meta-RL model had 10 parameter-value sets for the 11 participants and, lastly, there was only 1 RL best fitting combination of parameter values for the four participants described by this model. This demonstrates that even within the four groups of participants fit by the different models there are notable individual differences to be captured.

Across all participants that fit LTM containing models (which includes the integrated models), the memory decay parameter (*d*) had a negative skew: 85% of subjects fell on the highest level in the range explored and the next 11% were characterized by the second highest value of decay rate. For the retrieval noise (*s*) parameter on the other hand, fit parameters were, relatively more uniformly sampled from the range of values explored with a slight bimodal distribution (34% of fit participants had the highest noise and 20% had the lowest value). The last of the LTM parameters, the spreading-activation parameter, attempts to capture individual differences in attention and working memory (Lovett et al., 2000). The default level of 1 (in ACT-R 7x) and above injected too much spreading activation to capture individual differences so a range between 0.1 and 0.5 (Table 1) was used. Subject-fit values were positively skewed, where 47% of participants fell onto the lowest value 0.1, followed closely by 33% for the second lowest parameter value, 0.2.

Regarding the two RL parameters, τ(RL noise) and α(learning rate), both parameters are slightly skewed but in opposing directions. The noise parameter is positively skewed (40% of fit participants falling on the lowest values of 0.1).

There is not sufficient variability in these data to estimate the effects of the parameters in isolation on learning outcomes using linear methods. But it can be taken as evidence that perhaps, single parameters in isolation may not have large driving effects on learning outcomes; instead, learning behavior might be better explained by the combined effect of all or a majority of the parameters, as explored in part 1 of our 2-step approach. For instance, in integrated models, higher levels of noise in the LTM portion of the model encourages the meta-learner to prefer the RL portion of the model when lower levels of noise and higher levels of RL learning rate occur. Relationships like these are difficult to capture in linear models. The β in the Meta-RL model is explored in detail in the next section.

The 6th parameter, learning bias β, was pre-defined in the Biased model (figure 5) and had the values 20%, 40%, 60% and 80% probability of RL use. We found that 9 out of 11 participants that fit this model used RL only 28.8% (M = 0.289, SEM = 0.035) of the time, adding to the growing evidence of general preference of a declarative strategy that we have uncovered. The image is slightly more complicated when considering the Meta-RL model. Recall that the Meta-RL model used the RL-based, production utility learning to select either the RL or LTM learning models. The bias outcome, here, was measured at the end of learning by taking the mean number of times the RL sub-model was deployed for the set-size 3 and set-size 6 conditions. This, as predicted, was influenced by the relative success of the sub-models as determined by the current values of their parameters (Figure 14). For example, looking at the model data only, significantly higher levels of RL were selected to learn the set-size 6 block as the value of the learning rate

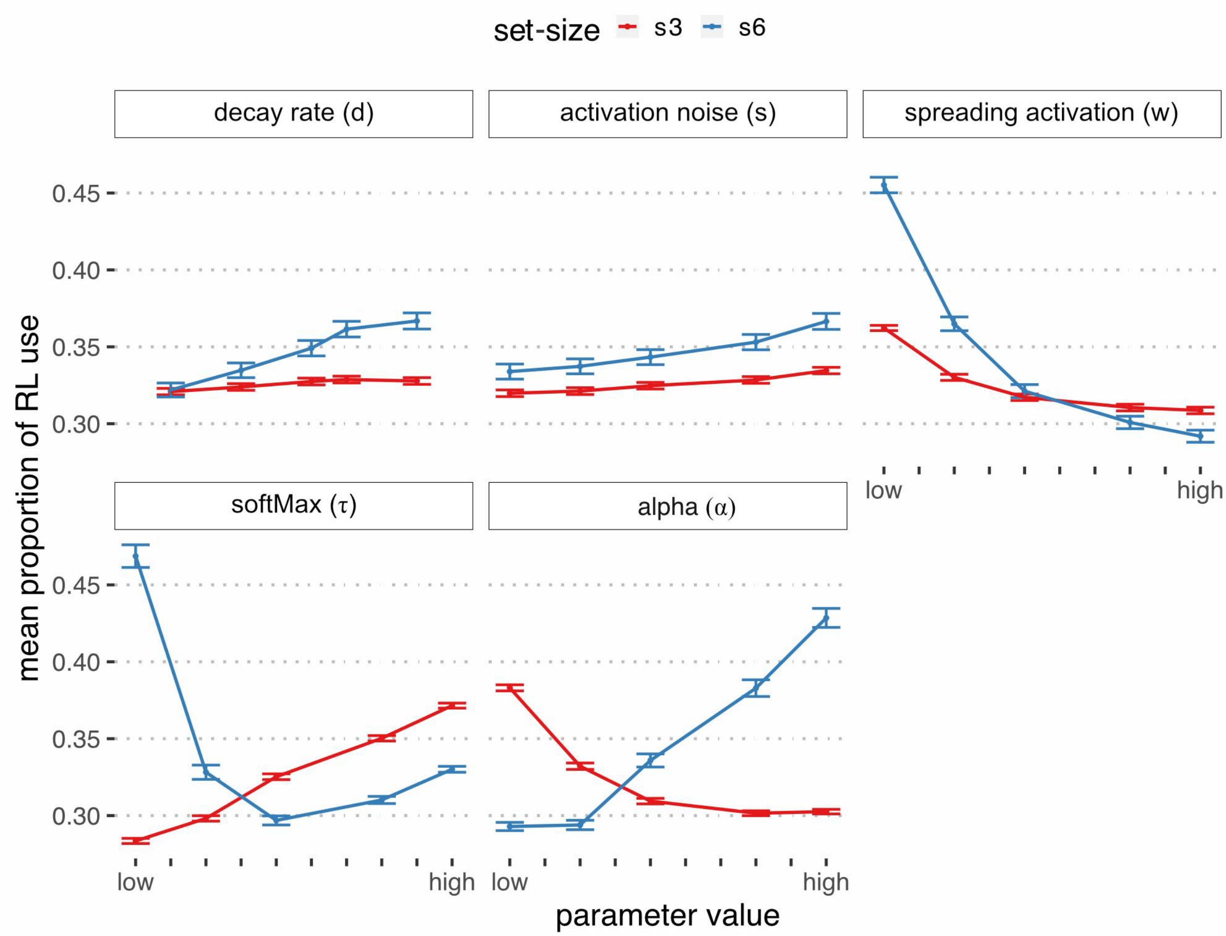


Figure 12: Model data only showing the effect of changes of parameter values on the proportion of RL used for the set-size 3 and set-size 6 blocks, in the Meta-RL integrated model. Top 3 panels are LTM parameters, and the bottom 2 panels are RL parameters.

alpha increased; however, a decrease in the proportion of RL used for the set-size 3 block was observed. Interestingly, the changes in the three LTM parameters exhibited similar trends in the proportion of RL used in both set-size blocks, but slightly more pronounced for set-size 6 blocks.

An increase in the noise and memory decay parameters resulted in more use of the RL subsystem. Similarly, an increase in the spreading-activation parameter, which favors the LTM model, led to a related decrease in RL use, again, at a more pronounced level for the set-size 6 block. Only 11 of our 83 participants fit this model best but we observed that the estimated bias towards RL in the set-size 6 was higher, around 45% of the time, compared to set-size 3 (set-size 3: M = 0.28; SEM = 0.05; set-size 6: M = 0.446, SEM = 0.045). RL was preferred, in general, to learn the set-size 6 blocks, which aligns with Collins' (2018) observation, but only a small subset of our participants used this distinct preference in strategies to engage with the two different set-sizes.

Discussion

The results reported herein extend our knowledge about the nature of stimulus-response learning in two important ways. First, we showed that long-term declarative learning strategies can explain patterns previously ascribed to shorter-term (or working) memory processes and reinforcement learning (e.g., Collins, 2018). Second, and perhaps more importantly, our data suggest that not all individuals use the same learning strategies to engage with even simple stimulus-response learning tasks. Below we discuss the implications of these results in greater detail.

The Costs and Benefits of an Idiographic Approach

The current study capitalized on the use of idiographic computational models — models designed to best fit a specific individual with a high degree of fidelity, rather than a group average. This approach has recently gained prominence in cognitive neuroscience (Ceballos, Stocco, & Prat, 2020; Daw, 2011). We used the ACT-R architecture to create four models that corresponded to different strategies that may be employed to learn stimulus-response mappings. When each individual's data was fit to all four model types, we observed that some learners adapt their strategy dynamically during the task (more likely to fit the meta-learning model), while others maintain a bias towards one learning system or the other (more likely to fit the Biased model). But *most* learners relied on an explicit declarative memory strategy, which in itself comes with a repertoire of domain-general, memory skills that learners likely took advantage of. Without an idiographic approach, elucidating these individual differences would not be possible.

The idiographic approach can also be used to estimate more stable latent variables or parameters that describe cognitive or neurological characteristics. These estimated parameters can be used to explain or to predict other individual or even group behaviors (Daw, 2011), circumvent experimental challenges like test-retest variability (e.g., Xu and Stocco, 2021), or even explain different aspects of seemingly disparate cognitive functions that are difficult to explain behaviorally, with unifying latent variables (e.g., Lovett et al., 2000, on working memory).

However, the idiographic approach comes with its own set of limitations. First, fitting models to individual behavioral data is challenging. Good fits are difficult to achieve due to the increased variability and noise in individual subject data. This procedure is also more costly in-terms of computation, especially when multiple models are compared. In our study for instance, we performed a grid search of constrained parameters by simulating data for each possible combination of parameters. This required several days of computation. And with increased granularity in parameter values, which is a future direction for capturing more nuanced individual differences, this time grows exponentially. A sparse, and narrow range, set of parameters sometimes results in under-, or over-estimating parameter values, blurring the boundaries between individuals, and reducing inter-subject variability. This loss in variability reduces our predictive power in relating our model parameters to other measures like EEG spectral power or resting state fMRI networks, which limits the reach of what we would be able to explain about learning. This was a specific limitation in our study that we hope to address in future studies.

Disambiguating Long-term Declarative Memory and Long-term Reinforcement learning

The success and prominence of RL theory in cognitive neuroscience may have resulted in an underestimation of how much individuals rely on declarative strategies, even when learning simple response association tasks. This is apparent in Collins' (2018) and Collins and Frank's (2012) conclusions, which, while acknowledging working memory, dismiss the possibility of participants forming long-term declarative associations altogether. Our results point to different conclusions — declarative long-term memory is a popular mechanism for most learners, even for this seemingly simple, feedback-based, stimulus-response-learning task.

The majority of our participants preferred a declarative memory strategy to learn both the set-size 3 and set-size 6 blocks (68.6% of participants fit the LTM only model best). This result suggests that at least some of the learning that Collins (2018) ascribed to RL, can also be explained by explicit, long-term declarative memory representation, which is driven by power laws of frequency and recency. Our RL only model fit only four participants. WM is another probable candidate mechanism for performing the RLWM task. But we argue that there was little use of maintained WM association during learning. For instance, for the set-size 3 objects, participants in the LTM group, on average, have lower accuracy on the post-test but the memory decay between training and test, suggesting memory decay. However, this decay is not *large enough* (a decrease of 13% in accuracy from training to test) to warrant a significant proportion of WM use. Similarly, the high count in the number of participants fit for the LTM model, over the RL model, signified that long-term declarative memory can also be used to learn difficult associations (set-size 6 blocks), much like RL, robustly. The smaller percentage in forgetting from learning to test (only about 9%) suggests even less use of WM during set-size 6 learning. It might be overextending our findings to suggest that no, or minimal, RL was used in the set-size 6 blocks, so future work intends to fit trial-by-trial data to further isolate and quantify RL and declarative LTM contributions during learning.

A preference for a declarative memory strategy is reflected in our hybrid models as well. Of The participants that fit the Biased model, 81% fit model simulations that utilized LTM 71% of the time, over the RL learning mechanism during learning. In the Meta-RL model, where RL versus declarative LTM contributions were estimated for each set-size condition and participant, we found that learners fit model simulations that preferred declarative LTM 72.6% of the time for set-size 3 and 55.6% of the time for set-size 6 trials. Therefore, learning in the current task likely also occurs through declarative memory with robust effects, even when RL is available as a viable mechanism. This is consistent with the increasing popularity of declarative memory-based approaches to learning and decision-making, such as the popular decision-by sampling (Stewart, Chater, & Brown, 2006), and Instance-Based Learning (Gonzalez, Lerch, & Lebiere, 2003) approaches.

The popularity of declarative memory is hardly surprising given the large set of mnemonic strategies for learning. Semantic associations, repeated exposure to stimuli (e.g., Anderson, 2000), and even the simple act of naming objects (e.g., Lupyan et al., 2007) results in better memory for associated responses or objects, all of which are strategies that rely on explicit declarative representations. Which specific declarative memory strategy is used for learning, or even preferring an RL strategy, is different for each subject. These choices result from individuals' learning history and expertise (e.g., Cetron et al., 2020), cultural background (e.g., Gibson et al., 2017) and stimulus complexity (stimuli with complex or too many features are difficult to learn using declarative mechanisms, e.g., Zeithamova and Maddox, 2006). Future research plans to model how stimuli complexity and learners' shared world knowledge might affect learning and strategy choices.

Individual differences in Strategy selection

Acquiring complex skills require a mixture of memory mechanisms (e.g., Anderson 2007; Anderson et al., 2021). For instance, solving an algebra problem requires retrieval of declarative mathematical facts and algorithms for solving a mathematical problem. Some of these may be new declarative knowledge from instructions, and other algorithms that were likely discovered by the learner through RL-based trial and error and are now proceduralized. Similarly, the RLWM task, though much simpler than algebra problems, is not so constrained in its design that multiple learning mechanisms cannot be employed (with the exception perhaps that set-size 6 blocks would be difficult to perform robustly through WM alone). From here, it is not a significant leap to posit that different learners would depend on their individual learning experience or meta-cognition and elect to use strategies that work for them.

We hypothesized that interaction between task contents and individual learning preferences might necessitate dynamic deployment of learning mechanisms. On-going appraisal of recent learning success and failure is one way a learner can adjust strategies to improve outcomes. This hypothesis is different from Collins' (2018) expectation that puts RL as the likely sole mechanism for learning difficult tasks with robust returns and WM for easier tasks, rigidly. Our Meta-RL model, on the other hand, employed an RL meta-learner which was designed to capture the likely situation where strategies are adapted depending on task demands and success rate. We found that the participants who fit the Meta-RL model used an RL strategy only at variable rates to learn the set-size 3 and set-size 6 objects, which aligns with our learning flexibility and adjustment hypothesis. But it also suggests that a mixture of learning mechanisms is often used even in seemingly simple learning tasks like the current task. Additional individual factors that might affect learning mechanism selection are discussed below.

Individual differences in learning may also arise due differences in sustained attention, working memory capacity, and declarative memory forgetting rate, among others. Our modeling paradigm has the potential to capture these aspects of human cognition that affect strategy selection and learning outcomes. For example, the RL meta-learner in the Meta-RL model, recruited either the LTM or RL sub-model depending on their relative success with learning. This selection process was heavily influenced by the combination of parameter values for RL learning rate and noise, LTM decay rate and noise, and the working memory equivalent ACT-R parameter spreading activation. In one instance, the proportion of RL use fell for both set-sizes as the spreading activation value in the LTM model increased, which made LTM more successful. In a second example, RL was more preferred for the set-size 6 block as RL learning rate increased but de-emphasized for the set-size 3 block and when learning rate was low. We can, therefore, venture to assume that individual learners similarly have varying intrinsic cognitive properties (WM capacity or memory decay rate) that affect learning strategy choice dynamically. To the best of our knowledge, this is the first study to report such findings.

Conclusions

This work highlights that individual differences in learning are present and prevalent, even in simple stimulus-response mapping tasks. As our results highlight, group-averaged data might not reflect the true behavior of any of its component individuals. Computational models provide a new and unique method to understand, measure, and uncover the dimensions in which individuals differ from one another. We have demonstrated here that individual learning behavior can be explained by different combinations of WM, LTM, RL learning strategies, which sometimes interact with task properties (different set-sizes, and image types, for example). For this reason, we advocate for developing idiographic (i.e., individual level) models within an integrated cognitive architecture, so that the different models benefit from a common, well-established set of constraints (as do Laird, Lebiere, & Rosenbloom, 2017).

References

Anderson, J. R. (2000). *Learning and memory: An integrated approach*. John Wiley & Sons Inc.

Anderson, J. R. (2007). How can the human mind occur in the physical universe? *Oxford University Press.*

Atkinson, R.C.; Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In Spence, K.W.; Spence, J.T. (eds.). The psychology of learning and motivation. 2. *New York: Academic Press.* pp. 89–195.

Ceballos, J. M., Stocco, A., & Prat, C. S. (2020). The Role of Basal Ganglia Reinforcement Learning in Lexical Ambiguity Resolution. *Topics in Cognitive Science*, 12(1), 402-416.

Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V., & Kraemer, D. J. (2020). Using the force: STEM knowledge and experience construct shared neural representations of engineering concepts. *NPJ science of learning*, *5*(1), 1-10.

Collins, A. G. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of cognitive neuroscience*, 30(10), 1422-1432.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7), 1024 1035.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, 25 (3), 315-353.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. Decision making, affect, and learning: Attention and performance XXIII, 23(1).

Fisher, C. R., Walsh, M. M., Blaha, L. M., Gunzelmann, G., & Veksler, B. (2016). Efficient parameter estimation of cognitive models for real-time performance monitoring and adaptive interfaces. *In Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016)*. University Park, PA.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences,* 104(41), 16311-16316.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785-10790.

Gluck, M. A. (2002). How do People Solve the "Weather Prediction" Task?: Individual Variability in Strategies for Probabilistic Category Learning. *Learning & Memory,* 9(6), 408–418.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591-635.

Hikosaka, O., Yamamoto, S., Yasuda, M., & Kim, H. F. (2013). Why skill matters. *Trends in Cognitive Sciences,* 17(9), 434–441.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122.

Kalra, P. B., Gabrieli, J. D., & Finn, A. S. (2019). Evidence of stable individual differences in implicit learning. *Cognition*, *190*, 199-211

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of experimental psychology: General*, 130(2), 169.

Kass, R. E., & Raftery, A. E., (1995) Bayes Factors. *Journal of the American Statistical Association*:90(430), 773-795.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, *38*(4), 13-26.

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research*, 1(2), 99-118.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological science*, *18*(12), 1077-1083.

McDonald, R. J., & Hong, N. S. (2013). How does a specific learning and memory system in the mammalian brain gain control of behavior?: Memory Systems and Behavioral Control. *Hippocampus*, 23(11), 1084–1102.

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, 100(2), 463-475.

Niv, Y. (2009). Reinforcement learning in the brain. J*ournal of Mathematical Psychology*, 53(3), 139-154

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.

Poldrack, Russell A, & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia* 41(2003) 245–2517.

Puig, M. V., Antzoulatos, E. G., & Miller, E. K. (2014). Prefrontal dopamine in associative learning and memory. *Neuroscience,* 282 (2014), 217–229.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171-177.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive psychology*, *53*(1), 1-26.

Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological review*, 117(2), 541.

Xu, Y., Stocco, A., (2021). Recovering Reliable Idiographic Biological Parameters from Noisy Behavioral Data: the Case of Basal Ganglia Indices in the Probabilistic Selection Task. *Comput Brain Behav* **4,** 318–334.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & cognition*, *34*(2), 387-398.

Acknowledgement

This work was supported by a grant from the Office of Naval Research (N00014-20-1-2393) awarded to Dr. Chantel S. Prat from the Department of Psychology at the University of Washington.

Chapter 3: Illuminating Individual Learning Dynamics Within a Task: A Computational Model Analysis

Theodros Haile (theodros@uw.edu)

Chantel Prat (csprat@uw.edu)

Andrea Stocco (stocco@uw.edu)

Department of Psychology, University of Washington, Seattle

Abstract

Individual learners rely on different strategies (e.g., different combinations of declarative and reinforcement learning) to acquire new skills, and also engage metacognition which might affect which strategies they rely on. In this study, we sought to use cognitive models to test if strategy changes occur over time and in response to changing task demands. To address those questions, we fit four idiographic ACT-R models to the first and second halves of a stimulus-response learning task (RLWM, Collins, 2018) in 83 participants. Additionally, we performed a parameter recovery procedure to test how well our model fitting procedure captures strategies and parameters. The parameter recovery procedure showed that while model types are confidently recovered, only the reinforcement learning model had parameters that were well recovered. In regard to the learning dynamics analysis, we found that the performance of 61 learners (73.4%) was best characterized by the same model in the first and second halves of the task. Comparing the two set-size conditions, we found that 51 participants (61%) fit the same models for set-size 3 and set-size 6. Of those, 41 participants fit the LTM model, aligning with our previous findings. Model fits did not align with the set-size 3 -LTM and set-size 6- RL expectations set by Collins, 2018, in fact we found no predictable differences in strategy use that was driven solely by change in set-size.

Keywords: individual differences, learning, learning strategies, reinforcement learning, declarative memory, ACT-R.

Introduction

Studies have shown that individuals approach learning tasks with different strategies (e.g., Gluck, 2002). While learning is a parallel process of dual memory systems — declarative and procedural memory— information representation and guidance of behavior, shift to procedural memory, after sufficient practice (Squire, 2004; Hill & Schneider, 2006; McDonald & Hong, 2013). Skill performance that is mostly executed through the procedural memory system bears the clear hallmarks of behavior that we associate with expertise – fast, nearly automatic behavioral responses to pertinent stimuli (e.g., Tenison et al., 2016). But we hypothesize that before a level of expertise is reached, a learner must engage an appropriate memory system during the initial stages of learning, as information in different tasks might be better represented by either declarative or procedural memory (e.g., DeCaro et al., 2008). In the current study, we seek to understand if this strategy selection process is dynamic or static in an individual. Additionally, if strategies seem to change during learning, we seek to test if there are advantages for doing so, reflected in learning outcomes.

In previous work, we identified learning strategies, using idiographic computational models, but the question of dynamics arose given that learners were engaged in learning for around forty minutes on a task that included different difficulty levels. Therefore, a more complete understanding of individual learning requires that we answer that question. In the current study, we hypothesize that learning success might crucially relate to not only individual differences in applying an appropriate learning strategy, but also to *finding* that appropriate strategy. This is metacognitive in nature, and is also influenced by cognitive characteristics (e.g., Stocco et al., 2021), discussed briefly below.

Metacognition and Strategy Selection

Metacognition or memory monitoring is a multi-stage processing that occurs at both stimulus presentation and recall ((McDonough et al., 2021). There are many factors that affect individual learning approaches, such as working memory capacity for example, which is related with better reasoning ability (Kane et al., 2005) and low distractibility (Unsworth et al., 2004). Both of these allow for better metacognitive monitoring, which learners have been known to engage (McDonough et al., 2021; Kelemen et al., 2000), and lead to strategy adjustments (e.g., Stocco et al., 2021). Individual differences studies have shown that learners who engage metacognitive strategies learn better (Stocco et al., 2021), but sometimes learners might poorly judge their own learning and lead to unwanted effects (Metcalfe et al., 2007; Kelemen et al., 2000). Regardless, interventions that include or increase meta-cognitive strategies lead to task improvement (Metcalfe et al., 2007). Given the impact that metacognition has on learning, this study seeks to extend the findings of our previous work by utilizing new analyses on the same data to find evidence of strategy adjustments and associated advantages during learning.

The relationship between strategy and metacognition is connected through individual differences in cognitive characteristics, as stated above, but current task demands might present additional constraints on which strategies can be used at a given time. This was, after all, part of the design of the task we used. Briefly, the RLWM task (Collins, 2018) involves learning associations between images and letters in two learning conditions, easy short blocks, and difficult long blocks for a total of 40 minutes. Collins' (2018) hypothesis stated that learning the long difficult block should engage RL mechanisms and the short blocks should engage WM. Collins (2018) fit group data, and all task conditions together, to a set of models and explained behavior with combined RL-WM model. Our previous work, while fitting individuals separately, still sought a single model that explained all behavior in all task conditions. We aim to take a different model fitting approach to examine how the interaction between task demands and individual cognitive characteristics results in different strategies.

We predict that two things might result here. First, as Collins (2018), forwarded, we might find that the set-size 3 conditions might be more frequently fit by the LTM or LTM-skewing models, given that this easy condition might be quickly learned that way. The set-size 6 condition on the other hand might be fit more frequently by RL or RL-skewing models.

Our second prediction simply states that we find, just like in Chapter 2, that LTM model fits most people for both set-sizes. This prediction stems from the fact that the RLWM task can be performed with either strategy, given the right conditions and that explicit strategies might be popular among our study participants, who are all college students. Learning in school occurs mostly through declarative instruction and explicit examples, and learning is assessed through recall or recognition tests. Additionally, students acquire meta-learning strategies, implicitly, for navigating these learning environments with the aim of maximizing academic success (Puro & Bloome, 1987).

Current Research Goals

In this study, we aim to capture learning dynamics using ACT-R (Anderson, 2007) based idiographic computational models as in Haile et al. (2020), on the RLWM task. Specifically, by fitting the two halves of the long learning phase, as well as fitting set-size 3 and set-size 6, separately, we hope to answer the following questions about dynamic strategy use during learning: (1) Do individuals switch learning strategies as they gain experience with a stimulus-response task? (2) If so, are there learning gains associated with switching tasks versus sticking with the original strategy deployed? (3) Are individual differences in switching related or driven by task demands (i.e., set-size 3 vs set-size 6)?

Capturing individual learning dynamics is important, but it will be just as informative to learn if strategy changes result in improved learning, and which specific changes lead to those improvements. As described above, the RLWM task has a 10-minute testing phase consisting of images from both block types and both halves of the task. We plan to group the test stimuli by which half and block they appeared in, so we can associate testing accuracy with the type of model that best fits it.

As a starting point however, we aim to obtain overall model fitting reliability by performing a parameter recovery procedure (e.g., Wilson and Collins, 2019). This step is necessary to establish the sensitivity and reliability of our models and model fitting procedure at identifying a learning strategy. Here, we will generate simulated 'participants' for each of our four models: LTM only, RL only, and the two integrated models, Biased and Meta-RL. These simulated participants will cover the entire field of parameter values so we can also test how well we can recover specific parameter values. The procedure is described below, but unlike the simulations for the testing models, which were averages of 100 simulations, these simulated subjects will have only the same numbers of blocks as human subjects (six for set-size 6 and 8 for set-size 3).

Materials and Methods

Participants

83 undergraduate students from the University of Washington participated in this experiment. All participants were monolingual English speakers recruited through the UW Psychology subject pool (47 females, aged 18-35 years) who received extra credit for their participation. Data were collected after receiving informed consent in one 2-hour session.

Behavioral Task

The Reinforcement Learning Working Memory task (RLWM, Collins, 2018) involves learning stimulus-response associations through a series of 14 blocks. Participants are instructed to respond with a keypress of either ‘C’, ‘V’ or ‘B’ to the displayed images. In 8 of the 14 blocks, participants learn to associate keypresses with three unique images, presented 12 times in random order. In the remaining 6 of the 14 blocks, participants learn to associate 6 unique images each presented 12 times within the block with the key presses stated above. The stimulus-response associations are deterministic, and participants learn through reward (+1 point for correct responses and 0 points for incorrect responses). The same sequence of set-size 3 and 6 blocks was presented to all participants. Following this learning phase, a 10-minute distractor task is administered before a surprise 206-trial test block. Participants make responses without feedback to items taken from both 3- and 6-set learning blocks. Stimulus presentations and data collection were done in MATLAB (mathworks.com) and Psychophysics Toolbox (Brainard, 1997).

Models

We built a series of four models in the ACT-R cognitive architecture to capture different learning strategies in the RLWM task (Anderson, 2007). We hypothesized that learners might use a single-component strategy, based on either the declarative long-term (LTM) or Reinforcement Learning (RL) memory systems or a multi-system approach with specific designs on integration. ACT-R was the optimal choice because of its integrated and flexible architecture for knowledge representation. ACT-R represents declarative memories as static records of information in its declarative module, and stimulus-response associations learned through reinforcement learning as conditional IF-THEN rules in its procedural module. These two modules interact with each other as well as with other perceptual and motor modules, capturing multiple aspects of cognition in a single framework. The use and acquisition of declarative and procedural representations are governed by a formal system of equations that capture the hallmarks of declarative and procedural memories, like memory decay over time, learning rate in response to feedback and, for explicit memories, the role of attention and working memory resources.

Declarative Learning (LTM) Model

This single-system model stores memories of specific task events, like stimulus images, responses, and outcomes, for later recall and use. If it has never encountered a particular stimulus before, it executes a random response, the outcome of which is stored for later recall. If it does have memory of a previous encounter, it attempts to retrieve a response that led to a correct outcome, it makes a random response otherwise. All attempts and outcomes are memorized.

In ACT-R, declarative memories consist of multiple identical *traces*, each of which decay over time according to a power function (Anderson, 2007; Anderson 2000). The availability of a memory m depends on its activation $A(m)$, which is the log function of the sum of its decaying traces. Activation can be momentarily increased through spreading activation, an attentional mechanism that can be used to maintain information for a brief amount of time and reflects the weights W given to any existing association between a contextual cues q and m. Formally:

$$A(m,t) = \sum_i (t-t_i)^{-d} + \sum_q WS_{q,m} \qquad (1)$$

We rely on three parameters that affect memory retrieval to capture individual differences: (a) activation noise s, which captures random fluctuations in a memory's activations and are associated with the probability of retrieval, (b) decay rate d, which captures the rate at which memories fade away and are forgotten (Sense et al., 2016); and (c) spreading activation weight W, which captures the attentional resources allocated, and has been shown to capture individual differences in working memory capacity (Lovett, et al., 2000; Daily et al, 2001).

Reinforcement Learning Model

This second single-system model uses production rules to represent all the possible stimulus-response associations in the RLWM task. The model initially responds randomly, until the correct rule accrues sufficient rewards to overcome the competitors as the task progresses, and the interface provides feedback. ACT-R's procedural module relies on reinforcement learning where the value or *utility* of a specific production, which contains a rule for a specific response, given a stimulus, is determined gradually through feedback.

We rely on two parameters that affect the utility of productions, learning rate (α), and selection noise (soft-max temperature (τ)). Specifically, each production rule p has an associated *utility* value, $U(p)$, that reflects its expected rewards and is learned through a temporal difference rule.

$$U_t(p) = U_{t-1}(p) + \alpha\,[R_t - U_{t-1}(p)] \qquad (2)$$

in which α is the learning rate and R_t is the reward given at time t. In our experiment, R_t is binary and corresponds to the feedback ("Correct", $R_t = 1$, and "Incorrect", $R_t = -1$) given by the task interface. Competing responses are selected on the bases of their respective utilities, using a soft-max rule controlled by a noise parameter τ.

Integrated RL-LTM Model: Biased

The simpler of our two integrated, multi-system models utilizes a bias parameter (β), in addition to the two RL and three LTM parameters. This parameter explicitly biases the model to use its LTM or RL sub-system to deploy to learn and respond to task trials. The bias is set in proportions of RL-use from mostly LTM at 20% to mostly RL at 80% in twenty percent increments. This model was designed with the expectation that learners might, somewhat rigidly, utilize a strategy that favors either RL or LTM or both, consistently throughout a learning task. The next model uses a more dynamic approach to address how systems might be integrated.

Integrated RL-LTM Model: Meta

This more complex version of our integrated model does not have additional parameters but includes meta-learning productions (prefer-RL and prefer-LTM) that are deployed dynamically throughout the task. They compete for task control through reinforcement learning, and the best subsystem, RL or LTM, is selected depending on how many rewards each was able to accumulate throughout the task. We measure what percentage of RL was used at the end of a simulation run. This model assumes that individuals are adaptive learners and can optimally choose strategies based on their relative success over a short time. For example, if the long-term memory strategy proves too difficult (as in the case of too many stimuli), the model would switch to a RL-based learning strategy. RL learned associations are shared with the LTM system by inserting explicit information into the memory module.

Experiment

As described above, we performed a parameter recovery analysis to test the reliability of our models and model fitting procedure, separately fit and compared models between the two halves of the task, and fit separately and compared models between the two set-size conditions.

Modeling Procedure

Each of the above models was run across a discretized range of its parameter space. Each model interacted with the same interface that displayed a stimulus, received response, and provided feedback. One simulation run contains 1 block of the short set-size 3 and 1 block of the longer set-size 6 condition. To obtain stable estimates, each model was run 100 times for each possible combination of parameters. In discretizing the range of each parameter, values were chosen to form an interval that surrounds the recommended value in the ACT-R documentation.

Model Fitting Procedure

Models were fit to each participant's data by selecting parameters that maximized each model's fit while penalizing the models' complexity. To this end, the Bayesian Information Criterion (Schwartz, 1978) was implemented as in Chapter 2, using the same linear approximation based on the residual sum of squares (see Results below).

Parameter recovery

To perform parameter recovery analysis, a new set of model simulations were generated with the intent to mimic a level of experience with the task similar to that of human participants. Recall that human participants encountered 8 set-size 3 blocks and 6 set-size 6 blocks. One run of a simulation is equivalent to just one block of learning on the two set-size conditions, so final simulated data were averages of 8 (set-size 3) and 6 (set-size 6) simulations. This produced considerably noisier simulations that closely resembled our human participants. These simulated "participants" were then fit with the original set of 100-run simulations.

Split-half analysis

In this analysis, we split the 14-block learning data in half (labeled Half-1 and Half-2). Each half contained equal numbers of set-size 3 and 6 blocks. The test phase of the task contained one large block of 210 images sampled from all learning blocks, so the images were filtered by their occurrence in Half-1 and Half-2 to separately measure test accuracy for each half. Finally, the halves were separately fit to models to identify learning strategies.

Split set-size analysis

In this analysis, we fit each of the two set-sizes, taking all of the data, separately. The goal of analysis is to find if behavior in the two conditions is explained by different strategies, above and beyond what we have attempted in our previous work (Haile et al., 2020). Since the difference in set-size is a significant part of the experiment design , we want to learn if task demands exert constraints on what learning mechanisms can be used despite individual dispositions. After all, as per Collins (2018), the two set-sizes were designed to be learned through WM for set-size 3 and RL for set-size 6. Our previous findings notwithstanding, a sound and thorough inquiry requires that we perform this analysis.

Results

Parameter recovery

In this two-pronged approach, we first performed the parameter recovery analysis. Here, as described above, simulated participants were generated using our ACT-R models and fit to 100-run simulations. All simulated subjects were fit to all four models and the best fit model was identified by selecting the fit that had the lowest BIC value produced by equation 3.

$$\text{BIC} = n + n \log(2\pi) + n \log(\text{RSS}/n) + \log(n)(k + 1) \quad (3)$$

Figure 1 Confusion matrix percent recovered strategy models

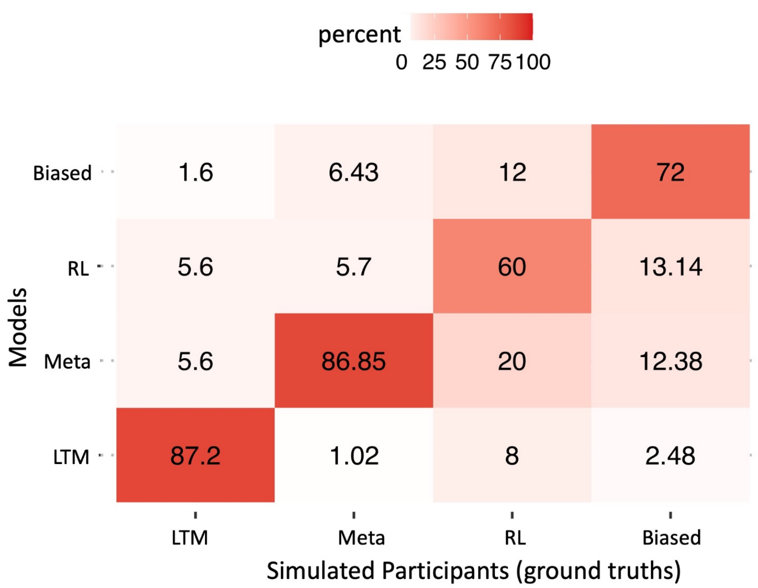


Figure 1: Proportions of simulated participants for each strategy type (x-axis, 8 simulations for set-size 3 and 6 simulations for set-size 6) correctly identified by the model (100 simulations per model). Values out of the diagonal show the proportions of simulated participants that were misidentified by the models.

We performed two sets of analyses that answered the questions: (1) Does the procedure correctly identify which model produced the data? And (2) Does the procedure identify which sets of parameters, for a given model, produced the data? Regarding the first question, we found that our models correctly identified where the simulated data came from 76.51% of the time. This percentage was highest for the LTM model at 87.2%, and lowest for the RL model at 60% (Meta: 86.85%; Biased 72%). Interestingly, 20% of the RL simulated data were identified as Meta (Figure 1). Next, we tested how well parameters were recovered. Here, our success rate was different for the types of models. We achieved high levels of parameter recovery for RL parameters but not for LTM parameters (Figure 2). Parameters for RL simulations were recovered well at $r = 0.76$ for α (learning rate) and $r = 0.79$ for τ (soft-max selection noise). For the LTM model however, recovered parameters for the LTM decay rate (*d)*, selection noise (*s)*, and spreading activation (*w),* were poorly correlated with simulation parameter values (*d*: $r = 0.22$; *s*: $r = 0.21$; *w* : $r = 0.35$).

Parameter recovery for the two mixed models showed similar patterns to that of the two single-system models but correlations were considerably lower for the Biased model, even for the RL portions of the models (α: $r = 0.22$, τ: $r = 0.38$, *w*: $r = 0.15$, *d*: $r = 0.18$, *s*: $r = 0.03$), but well recovered for most parameters in the Meta model (α: $r = 0.51$, τ: $r = 0.69$, *w*: $r = 0.53$, *d*: $r = 0.44$, *s*: $r = 0.10$). However, the bias parameter (*β)* was better recovered for the Biased model ($r = 0.58$) than the estimated bias in the Meta model ($r = 0.40$). This comparison takes advantage of the single-system models by providing a frame of reference, but it should be stressed that the parameters affect model performance in unison. For instance, when looking at the Meta model alone, preference for RL or LTM sub-systems (estimated post-hoc) was influenced by all the constituent parameters together

Figure 2: Correlation between recovered and ground truth parameter values

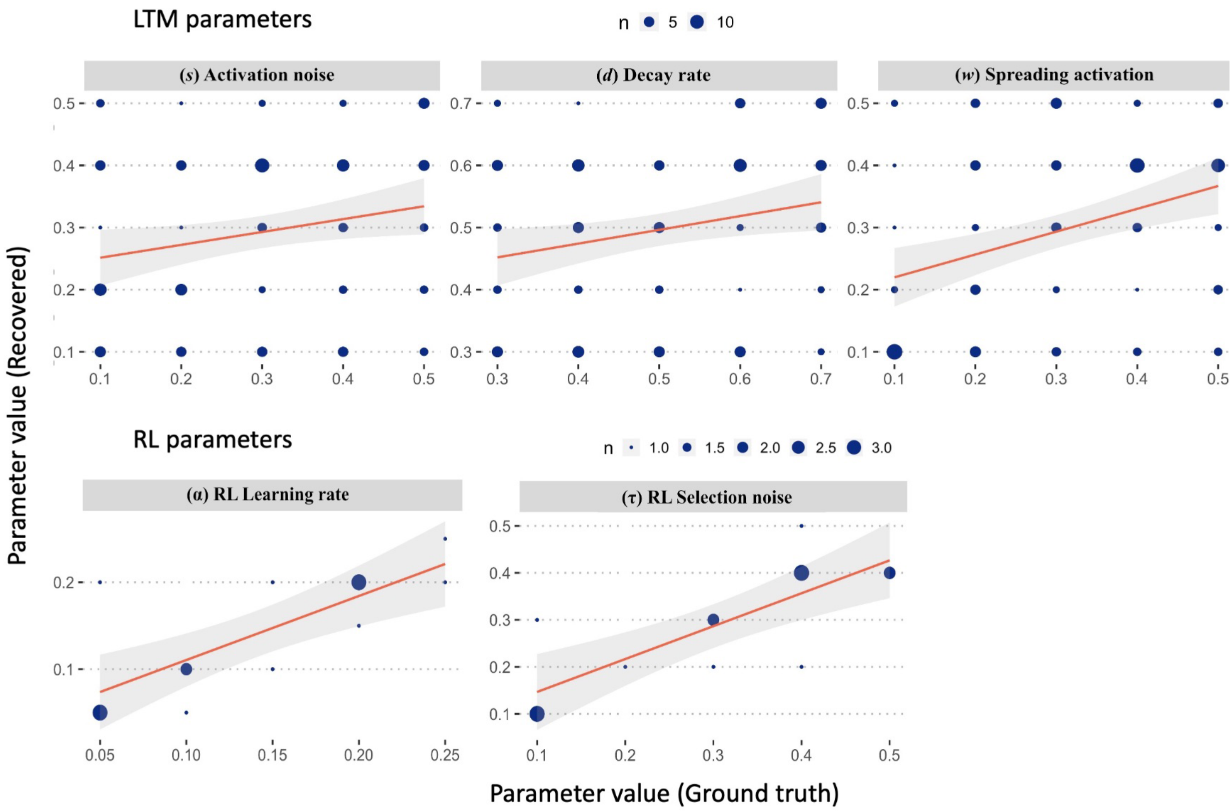


Figure 2: Correlation of recovered (y-axis) RL (bottom row) and LTM (top row) parameters (y-axis) with true parameter values from ground truth (also known as simulated participants) simulations (x-axis). Size of the markers shows the count of overlapping points. For the RL parameters RL learning rate and selection noise only the 15 unique parameter-sets, out of 25, that were correctly recovered as RL are shown. For the LTM parameters in the top row shown are the 109 unique parameter-set simulations (out of 125) that were correctly identified as coming from the LTM model.

Split-half analysis

Next, we sought to test if participants relied upon the same strategies to learn the associations throughout the task. So, the learning data were split into the first and last 7 blocks to compare learning outcomes and fit to models separately to identify strategies that might have led to those learning outcomes.

We first correlated behavioral learning outcomes - learning accuracy and test accuracy- for the two set-size conditions, from Half-1 to Half-2 (Figure 3). This analysis was done to ensure that participants behaved similarly across the two halves of the task. Correlation was highest for learning accuracy in set-size 6 condition ($r = 0.67$, $p < 0.01$), and lowest for learning accuracy in the set-size 3 condition ($r = 0.34$, $p < 0.01$). Correlations between the two halves for the testing conditions were high and significant (set-size 3: $r = 0.57$; set-size 6: $r = 0.65$; $p < 0.01$). Additionally, there is not a significant main effect of blocks in the 2 halves ($F(1, 656) = 0.96$, $p = 0.33$). This suggests that performance in the two halves, for most participants, is stable, and perhaps likely reliant on the same learning strategies, but we sought to answer this next, more robustly, by fitting models to each half and comparing identified learning strategies.

Figure 3: Correlation of behavioral accuracy scores for set-size 3 and 6 between Half 1 and Half 2

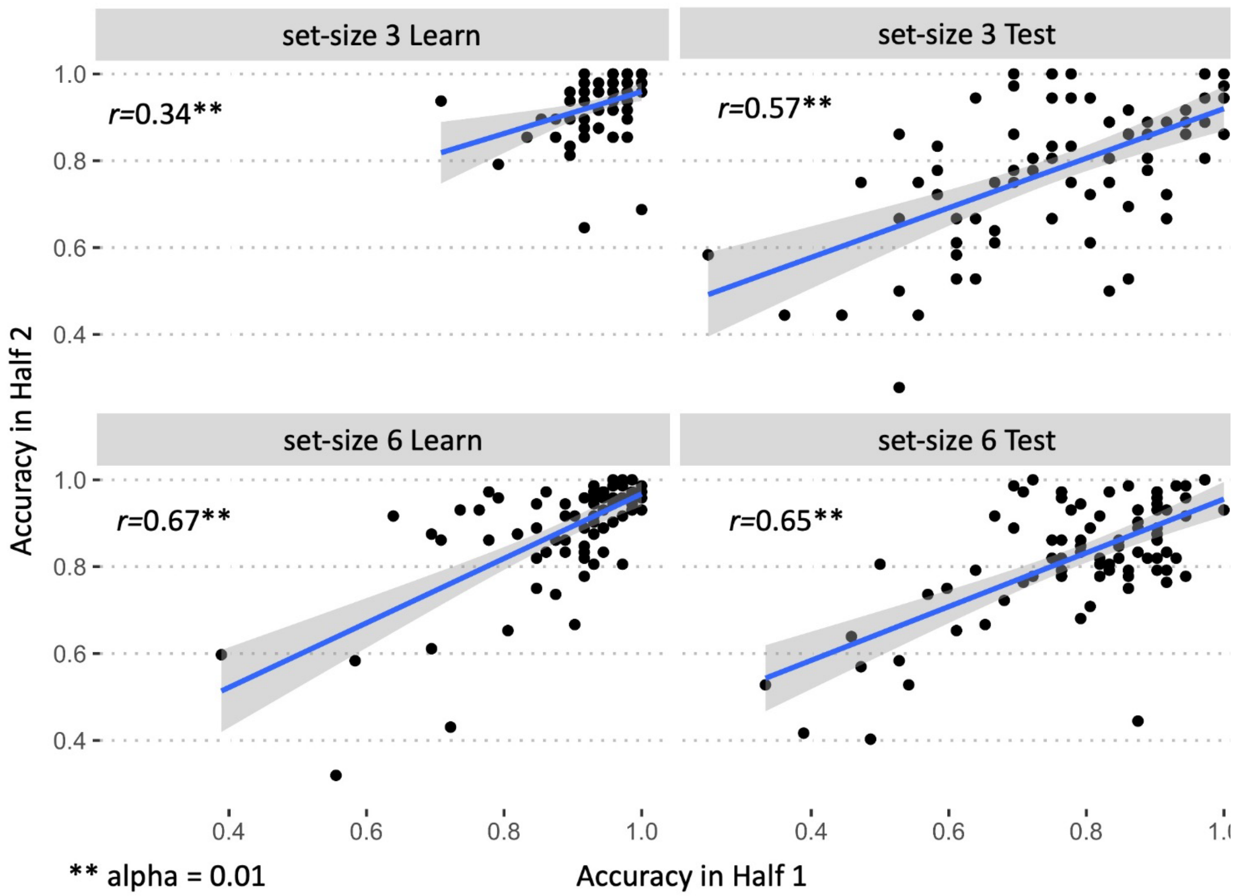


Figure 3: Accuracy scores from the end of learning (last 3 trials, left column) and testing (right column) phases were correlated for set-size 3 (top row) and set-size 6 (bottom row), across the Half 1(x-axis) and Half 3 (y-axis). Correlations between the two halves are significant for all conditions at alpha 0.01. A ceiling effect is evident in learning accuracy score for set-size 3.

After fitting the models to each half separately, model predictions were compared. We found that 73.4% (n = 61) of participants fit the same models in the two halves, suggesting that strategy use was stable throughout the learning task. This percentage is higher (86%) for participants who fit the LTM model in Half-1 (Figure 4). It is important to note here that none of our participants fit the Meta model in either half and that most participants fit the LTM model most (n = 56, in Half-1), and the RL model least (n = 11; Biased: n = 16, in Half-1).

Figure 4: Confusion matrix showing percentage of participant model fits across the two halves

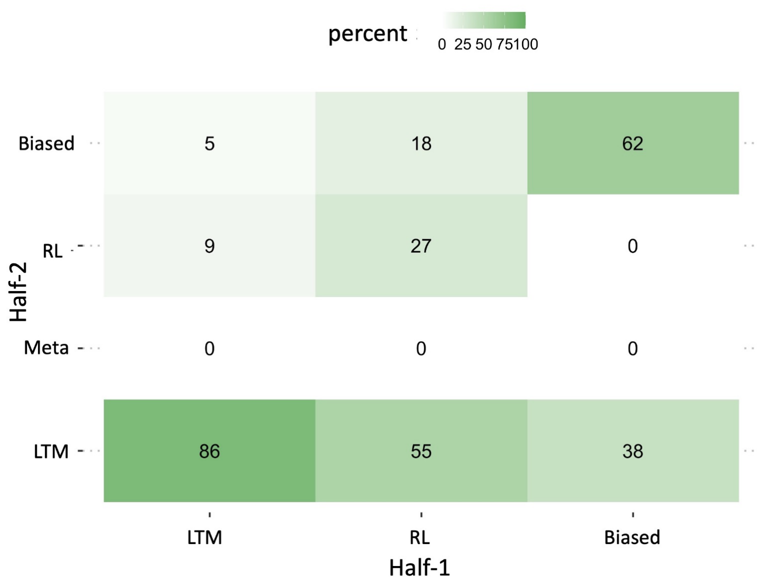


Figure 4: Matrix showing distributions of the model fits across the two halves based on where participants landed in Half-1 (columns add up to 100%).The diagonal shows proportions of participants who fit the same models in both halves. Out of diagonal percentages represent those participants who fit different models in Half-2, based on which model fit them in Half-1 (x-axis).

Figure 5 :Change in testing performance between half-1 and half-2.

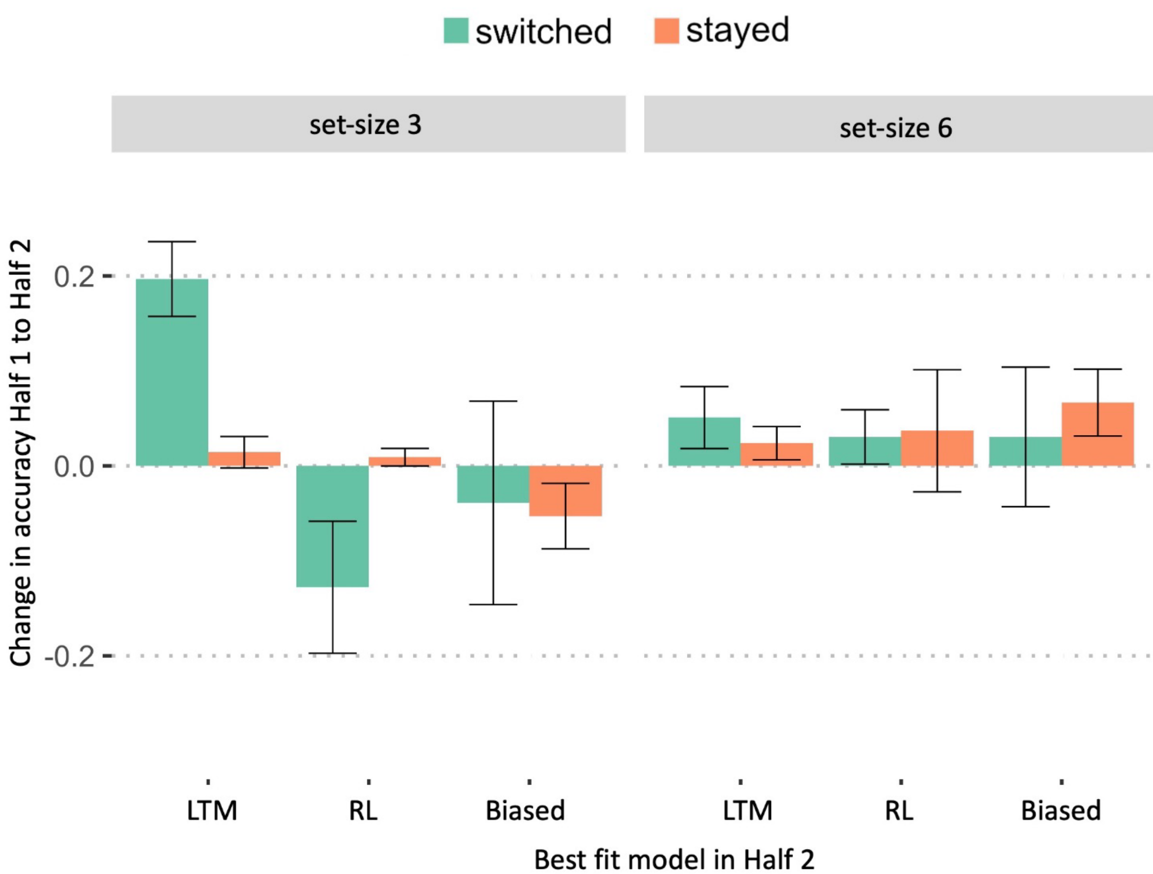


Figure 5 :Shown are the mean differences in accuracy (y-axis) between Half-1 and Half-2 for the testing condition. The x-axis shows participants grouped by the best fitting model in Half-2 since this analysis represents which strategy is ultimately useful for performance. "switched" group represents participants who fit different models in the two halves; "stayed" represents those who fit the same model in the two halves. Of the 22 participants who fit different models in the two halves, 12 fit LTM, 5 fit RL and the remaining 5 fit the Biased model.

Let us assume for a moment that we were able to identify a true switch in strategy for the minority of participants who fit different models in the two halves (n = 22), is there a measurable benefit in learning outcomes for this group? Our results indicate that there were no statistically significant differences between the two halves, in-terms of accuracy, during learning and test phases within the two set-size conditions, when comparisons were agnostic to best-fit model types. However, when split up by best-fit model type, "switching" to a LTM strategy was associated with a higher increase in accuracy from learning to test, by almost 20%, but *only* for set-size 3, and *only* during test (Figure 5B). Similar benefits were not observed for the set-size 6 testing phase, and both set-sizes for accuracy at the end of learning. Accuracy, however, tends to be lower in the second half for the learning phase (Figure 5A).

Split set-size analysis

In this analysis, we fit models to each set-size separately following the same procedures for model selection as above. We expect that, as our previous results show, the LTM model might be the most popular model for most participants, for both set-sizes. It will also be interesting to see if our findings match the theories forwarded by Collins (2018): LTM (which Collins calls WM) should fit the set-size 3 condition, and RL should predominate the set-size 6 condition.

To start 61% of participants (n = 51) fit the same models for both set-sizes. Of these, 80% (n = 41) fit the LTM model for both set-sizes, 17.6 % (n = 9) fit the Biased model for both, and only 1 subject fit both for the RL model.

Of more interest, perhaps, is the minority group that fit different models for the two set-sizes (39%, n = 32, Figure 6). Here, 50% fit the LTM model for set-size 3, and the remaining 50% were split evenly (8 participants in each) between the Biased and RL models. As for set-size 6, only 1 participant fit the RL model, 15 out of the 32 fit the LTM model, while 16 of the 32 fit the Biased model. It is surprising here that the LTM model is not the clear majority, and it would seem that our findings do not align with Collins (2018) predictions (Figure 6).

There are two additional dimensions to consider here. Firstly, what is the nature of the division between LTM and RL in those participants who fit the Biased model? Secondly, what is the pattern of fits between the two set-sizes for those who fit different models?

Figure 6: Proportions of models in the two set-sizes for those who fit different models.

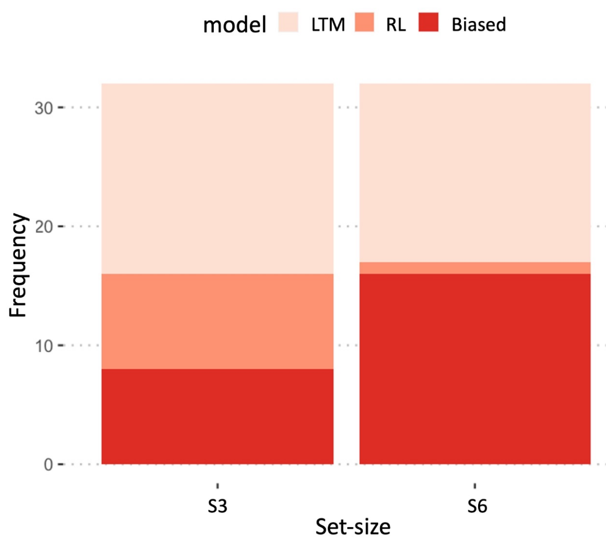


Figure 6:This plot shows the proportion of best fit models in "switchers", those who fit different models in the two set-size conditions. Fewer models represent participants in set-size 6 compared to set-size 3. β parameter in the Biased model (shown in red) skews more LTM in set-size 6 compared to set-size 3.

To answer the first question, we looked at the bias (β) parameter for those who fit the Biased parameter for each of the set-sizes. Recall that the β is a parameter that represents the percentage of RL use from 20 to 80% in 20% increments. We found that, for those who fit the Biased model for set-size 3, they skewed RL (β: M = 0.6, SEM = 0.093, n = 8). Similarly, those who fit the Biased model for set-size 6 also skewed RL but to a slightly lesser degree (β: M = 0.572, SEM = 0.062, n = 16). This paints a slightly different picture than we are used to, one that skews RL, compared to the overwhelming popularity of LTM that we have seen in our previous work and in the analyses before.

Secondly, if participants fit LTM or RL in set-size 3, they are more likely to fit LTM in set-size 6 (LTM: 71.93%; RL 88.9%). If participants fit the Biased model for set-size 3, a minority fit LTM in set-size 6 (41.8%), only one participant fit RL, and the rest, 52.9% fit the Biased model. Surprisingly, different levels of bias represent performance in set-size 3 and 6, for those who fit the Biased model for both set-sizes: (β: set-size 3: $M = 0.578$, $SEM = 0.085$; set-size 6: $M = 0.375$, $SEM = 0.07$). This is still a change from our expectations, especially from those predicted by Collins (2018), since it seems like that set-size 6 was characterized by more LTM use than set-size 3 and vice versa.

To supplement the results above and given that β values were different for participants in set-size 3 and 6, I represented all subjects on the same β scale and performed a Chi-squared test of independence. In this analysis, all participants who fit the LTM model were assigned a β value of 0, and all those who fit the RL model were assigned a value of 1. A Chi-squared test revealed that there are no significant differences between the two set-sizes across the β value frequencies (Figure 7): $X^2(5, 83) = 8.8$, $p = 0.115$.

Figure 7: Histogram of counts of participants on the β scale.

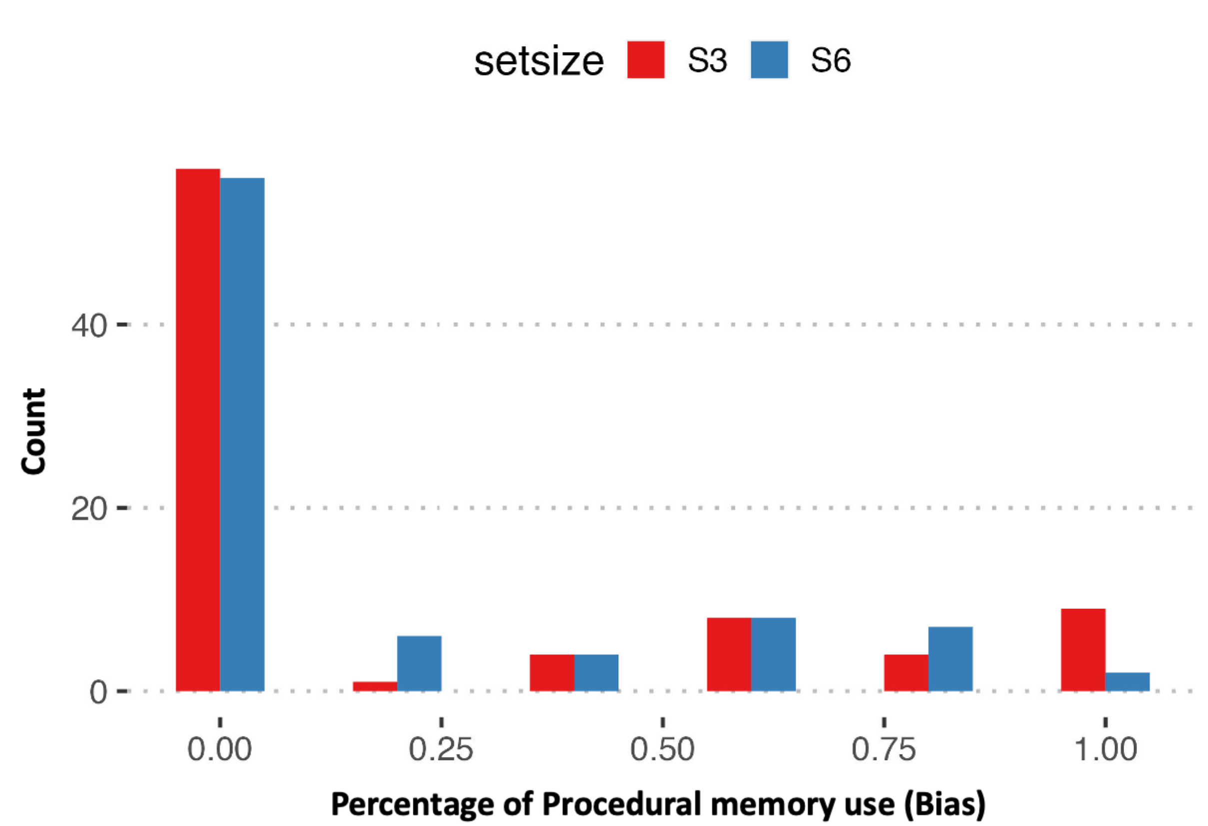


Figure 7: Participants across all model fits - LTM, RL, and Biased were represented on the same β (Bias) scale by assigning those who fit LTM a value of 0 and those who fit RL a value of 1. Most participants fit the LTM model for both set-sizes but, comparing frequencies of fits across the β scale revealed no significant effect of set-size on strategy.

Lastly, we ask a final important question: does a change in strategy mean better performance overall across the two set-sizes? To answer this question, we averaged across the two set-sizes for both learning and testing accuracy and compared those participants who fit the same models with those who fit different models. We found that mean accuracy was not affected at all by strategy type. This aligns with our previous findings that the RLWM task can be performed successfully by LTM, RL or a combination of both. That being said, it is important to learn why strategy changes occur in an individual, as other more realistic and complex tasks might not be so forgiving.

Discussion

How learners use their available memory resources, what we call a learning strategy, affects how well they acquire new associations or skills. We hypothesized that some learners might rely mostly on single memory mechanisms, like declarative long-term memory, or use a mixture of declarative and procedural memory when learning a task. But it is not clear if strategy selection at the individual level is stable. That is, once a learner lands on a strategy, do they use it for the duration of learning, or do they tend to alter their learning approach? Changes in strategy may perhaps be based on meta-cognitive evaluation of learning success, or differing task demands or changes in motivation. In this study, we addressed this question by breaking up a long stimulus-response learning task into two halves, which have identical blocks, and testing if different models explained learning in the two halves.

First, we performed a parameter recovery procedure to demonstrate that our four ACT-R models can capture which strategies likely led to specific patterns of behavior. This resulted in as high a congruence as 87% for the declarative only model (LTM), and as low as 60% for the reinforcement learning only model (RL). This suggests that the LTM strategy is highly recoverable, and therefore lends a level of certainty wherever we observe that a participant was best fit by this strategy.

However, success in recovery of the generating parameters was mixed. We were able to obtain high correlations between true and recovered parameters for the RL model but not for the LTM model. This is perhaps owing to the fact that we only have 2 parameters for the RL model - learning rate (α) and selection noise (τ), where pairs of values resulted in unique patterns of behavior. For instance, low performance is only evident in situations where there is high noise, low learning rate or the combination of the two. But the LTM model has three parameters, memory decay rate (*d)*, retrieval noise (*s)*, and spreading activation (*w)*, which leads to more ambiguous instances. In other words, many more combinations of parameter values could lead to similar results. This is a vital limitation. While it is important and encouraging that we can identify which strategy probably led to learning behavior, it is just as important to learn what set of cognitive characteristics led to them. The latter offers us insight into individual learning characteristics or aptitude that we can possibly extend to other tasks, which is the focus of the next chapter.

Proceeding with our planned analyses, we fit the models to each half of the learning data separately and compared model fits. We sought to understand learning dynamics because it might be one of the determining factors of successful learning outcomes. For instance, an individual with low memory decay rates might rely on their ability to make durable semantic association between stimulus-response pairs. As the task goes on, doing so might lead to more fatigue and ultimately poor performance in the second half of the task. We found that many participants (72%) fit the same models in the two halves. This was considerably higher for the participants who fit the LTM model in the first half. This suggests that most participants utilize LTM to start but also continue to use this strategy throughout the task. Most participants also fit the LTM model in the second half, regardless of which model fit them best in the first half, suggesting that LTM is ultimately the best choice for learning associations in this task.

It also appears that if a learner did not start out with a strategy that resembled LTM, they made a switch. We cannot yet identify why a strategy change occurred, but, surprisingly, this switch was associated with large gains, about 20%, in learning, at least for test accuracy in the set-size 3 condition. Switching to RL or the Biased strategy, actually resulted in losses for the set-size 3 condition. Furthermore, these gains were not observed in participants who did not switch strategies. There was minimal gain in the set-size 6 condition for switching to an LTM strategy and these were not different from the group that did not switch. There might be an interaction effect between the set-size condition and time spent on the task. The set-size 3 condition might allow for more flexibility in strategy use, but most choose LTM, while set-size 6 doesn't allow for a lot of flexibility, but strategy choice does not matter. Unequal and limited numbers of participants in each group, and complexity of doing a 2 (Half 1 vs Half 2) X 2 (set-size 3 vs set-size 6) X 3 (LTM vs RL vs Biased model) analysis precludes us from confirming this hypothesis.

In our third analysis we sought to identify if learners adopted different strategies for each of the set-sizes. Recall that the set-size condition in the RLWM task was designed to evoke different learning mechanisms. In our previous approach we fit both set-sizes together to demonstrate that a single strategy, influenced by specific parameter values might explain behavior in both set-sizes, as we have shown in Chapter 2. However, a more thorough analysis should involve individual model fits that will more clearly demonstrate if the same or different models explain behavior in the two set-sizes. One possibility here was that the set-size 6 condition would be fit mostly by RL, as was Collins' (2018) hypothesis.

About 61% of participants fit the same model for both set-sizes, and a large majority of those fit the LTM model. This aligns with the findings in Chapter 2 where a majority were explained by LTM even when fitting the two set-sizes together. For the minority who fit different models however, a complex pattern emerges. The set-size 3 condition was explained by LTM (the largest group, about half of participants), RL (about a quarter of participants), and an additional quarter, were best explained by 60% of RL use. This aligns with the theory above that set-size 3 allows for more freedom in strategy deployment. However, when we examine how β shifts across the two set-sizes for those who fit the Biased model, along with those who fit LTM and RL, there appear to be more dynamics that are not accounted for by just the change in set-size.

Unlike the above split-half analysis, a change in strategy is, surprisingly, not associated with a change in general learning outcomes, for both learning and testing (i.e., mean accuracy on the task regardless of set-size). There is no difference between those participants who fit different models and those who fit the same models. As mentioned above, our simulations demonstrate that the RLWM task can be performed successfully by either of the learning models/strategies — this might represent an extension of that finding.

There are limitations to our modeling effort that deserve mention. It can be argued that an explicit, declarative strategy is popular with our participants, perhaps because they are university students where a lot of learning is instructed, declarative, driven by explicitly described examples, and discussions amongst peers and instructors (Puro & Bloome, 1987; VanLehn, 1996). Therefore, it is not too far a leap to suggest that most students would rely on semantic links between stimuli in the same block. The stimuli in this task are also rich in detail and amenable to forming idiosyncratic semantic categories, which are supported by declarative memory. But our model fitting procedure might be biased towards the LTM model as it is also a relatively simple model with a small number of parameters. The BIC function that we minimize during model fitting penalizes larger models, so we tend to see fewer numbers of those models explaining learning behavior even when a mixed-mechanism strategy is perhaps closer to the truth (e.g., Poldrack et al., 2001; Gluck, 2002).

In conclusion, participants are likely to stick to the same strategy throughout the task, and an LTM strategy best describes behavior for most. Using an LTM strategy is also related to better performance in set-size 3, especially if it is a strategy one comes to in the second half. However, our pattern of results suggests that participants use a more diverse set of strategies to learn the set-size 3 condition than the set-size 6 condition. Ultimately, though, a mostly LTM strategy seems to suit most learners best. While we can label learning behavior with the model that likely produced it, we cannot confidently estimate the parameters that led to that behavior. This limits what we can explain about our learners and their learning outcomes. In addition to that, we explored a coarse-grained, narrow set of parameters for our models, which is an additional limitation. We hope to address these limitations in future studies by providing validation from independent tasks and exploring more fine-grained parameter sets.

References

Anderson, J. R. (2007). How can the human mind occur in the physical universe? *Oxford University Press*.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V., & Kraemer, D. J. (2020). Using the force: STEM knowledge and experience construct shared neural representations of engineering concepts. NPJ science of learning, 5(1), 1-10.

Collins, A. G. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of cognitive neuroscience, 30*(10), 1422-1432.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. Cognitive Science, 25 (3), 315-353.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*(1), 284-294.

Haile, T., Prat, C. S., & Stocco, A. (2020). One size doesn't fit all: Idiographic computational models reveal individual differences in learning and meta-learning strategies. In *Proceedings of the 18th International Conference on Cognitive Modeling*.

Heitzmann, N., Stadler, M., Richters, C., Radkowitsch, A., Schmidmaier, R., Weidenbusch, M., & Fischer, M. R. (2023). Learners' adjustment strategies following impasses in simulations-Effects of prior knowledge. *Learning and Instruction*, *83*, 101632.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. Psychological review, 99(1), 122.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005).

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & cognition*, *28*, 92-107.

Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition*, *169*, 84-90.

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. Cognitive Systems Research, 1(2), 99-118.

McDonough, I. M., Enam, T., Kraemer, K. R., Eakin, D. K., & Kim, M. (2021). Is there more to metamemory? An argument for two specialized monitoring abilities. *Psychonomic Bulletin & Review*, *28*(5), 1657-1667.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology*, *19*(4-5), 743-768.

Puro, P., & Bloome, D. (1987). Understanding classroom communication. *Theory into practice*, *26*(1), 26-31.

Schwartz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461– 464.

Shute, V. J. (1991). Who is Likely to Acquire Programming Skills? Journal of Educational Computing Research, 7(1), 1–24

Stocco A., Lebiere, C., & Anderson, J. R. (2010).Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological review*, 117(2), 541

Stocco, A., Prat, C. S., & Graham, L. K. (2021). Individual Differences in Reward-Based Learning Predict Fluid Reasoning Abilities. *Cognitive Science*, *45*(2), e12941.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. ELife, 8, e49547.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and individual differences*, *8*(4), 327-353.

Chapter 4: One-size does not fit all–but tailoring is difficult: Exploring the generalizability of model parameters to other cognitive tasks.

Theodros Haile

Chantel Prat

Andrea Stocco

Department of Psychology, University of Washington, Seattle

Abstract

This study aimed to assess the predictive power of estimated model parameters - that reflect latent cognitive characteristics - on individual strategy and learning success. Additionally, this study sought to validate parameter values derived from the RLWM task (Collins, 2018) and model by comparing them to values obtained from independent tasks. The ACT-R parameters (rate of forgetting (RoF, memory decay, Pavlik & Anderson, 2008) and spreading activation (*S*, working memory capacity analog) were estimated using the Adaptive Fact Learning Task (AFL, Van Rijn et al., 2009) and Modified Span Task (MODS, Daily et al., 2001), respectively. We found that parameters were good predictors of behavior with-in tasks but not across tasks. RoF and *S* estimates from RLWM task were not correlated to those obtained from AFL and MODS tasks, suggesting that parameter estimates were context-specific and lacked predictive power outside those contexts. We emphasize the importance of considering both cognitive characteristics and task demands in future research. Ultimately, parameters, while valuable, should be interpreted within specific model contexts, acknowledging the impact of design choices on their generalizability.

Keywords: Individual differences, ACT-R, declarative memory, procedural memory, idiographic model, rate of forgetting, working memory capacity.

Introduction

Learning a skill, or even simple associations is a process that involves the engagement of multiple memory mechanisms (e.g., Squire, 2004; Stocco et al., 2010; Anderson, 1982; Poldrack et al., 2001; Ashby and Crossley, 2010). Furthermore, cognitive characteristics like working memory capacity (e.g., Conway et al., 2005; Stocco et al., 2021; DeCaro et al., 2008), memory decay rates (Sense et al., 2016), and meta-cognitive strategies (e.g., Unsworth, 2016) that vary between individuals affect how quickly associations are acquired and how durable they are. We hypothesize that reliable measures of these cognitive characteristics might relate to the strategies (combination of declarative memory (LTM) and reinforcement learning, RL) learners engage in a specific environment. In the current study, we utilized model parameter estimates from cognitive models to measure cognitive characteristics as those have been shown to be more robust than behavioral measures alone (e.g., Xu & Stocco, 2021). Ultimately, we aimed to relate those individual parameters to learning behavior and strategy.

To date, our modeling explorations have investigated the way strategy deployment changes both between and within participants. In Chapter 2, we asked if cognitive architectures could serve as psychometric tools that reveal the different ways learners engage their memory resources. We showed that different models of memory system deployment, or what we have called strategy use, explained different learners' behaviors. In Chapter 3, we demonstrated that many learners deploy the same strategies across time, and when faced with different demands derived from varying levels of the same task. In the current study, we dig deeper, with the goal of exploring how the LTM, working memory and RL parameters estimated from a set of models in a variety of tasks relate to strategy and learning behavior. We further seek to use these cross-task, multi-model parameter estimates to test if parameters indicate generalizable cognitive characteristics.

In our previous modeling efforts, we found that the LTM only model best explained learning for the majority of learners. This suggests that many discriminable differences among learners in a simple association learning task might be explained by differences in parameters related to that model. However, the parameter recovery procedure we attempted in Chapter 3 (see Wilson and Collins, 2019) revealed that parameter values for the declarative memory model were relatively poorly recovered. Because of this, our ability to confidently state that the declarative memory parameter values that were associated with the best fitting model, reflect a true, replicable parameter.

Because the LTM only model has best explained behavior on our paired associate learning task, the Reinforcement Learning Working Memory (RLWM, Collins, 2018) task to date, the current exploration takes a two-pronged approach to attempt to improve our ability to capture reliable individual differences. First, I made several small changes to the declarative memory model (detailed in the approach section) which we believe better reflect the deployment of the cognitive processes parameterized by the model with the hope of better capturing reliable individual differences. Second, we sought to improve and validate our parameter fitting process by using independent tasks that have been shown to reliably estimate them. Taken together, the goal of the current experiment is to use a modified version of one of the models previously used to characterize performance on the RLWM task, along with two new tasks, the Modified Digit Span Task (Daily et al., 2001) and a paired association learning task called the Adaptive Fact Learning task (Sense et al., 2018, provided by MermoryLab LLC) for independent parameter estimation. In the section below, we describe the parameters we chose to target for this exploration in further detail.

Long Term Memory (LTM) Parameters of Interest

ACT-R's (Anderson, 2007) model of declarative long-term memory (LTM) has two parameters of interest that have been related to measurable cognitive characteristics that impact memory and task performance: rate of forgetting (called α, RoF is preferred here) which has been shown to reflect memory decay rate (Van Rijn et al., 2009; Pavlik & Anderson, 2008) and spreading activation which has been related to working memory capacity (Lovett et al. 2000; Daily et al., 2001). These parameters are part of two concepts that define the odds of successful memory activation (A_i in Equation 1): base-level activation (B_i in Equation 1) and spreading activation ($W_j S_{ji}$) which are summed together.

$$A_i = B_i + \sum_{j \in C} W_j S_{ji} \qquad (1)$$

Each of these parameters has been shown to capture individual differences in task performance (Van Rijn et al., 2009; Daily et al., 2001) and thus they will be the focus of the current investigation.

Rate of Forgetting (RoF).

The odds that any memory chunk will be successfully retrieved (i.e., if it has a high enough activation to reach a retrieval threshold) in ACT-R depends on that chunk's base-level activation (B_i in Equation 1 and Equation 2). The base-level activation in turn is determined by how many times it has been retrieved in the past (t_k in Equation 2) and, importantly, by its rate of decay (d in equations 2 and 3), which is most influenced by the RoF parameter (α, in Equation 3, Pavlik and Anderson, 2008; Van Rijn et al., 2009).

$$B_i = \ln\left(\sum_{k=1}^{n} t_k^{-d}\right) \quad (2)$$

Additionally, as dictated by the spacing effect, memory traces that are spaced further apart are better recalled than those occurring in closer succession, captured well by Pavlik and Anderson's (2008) model of the spacing effect (an important feature as this condition appears in our task and model). The second piece of this puzzle, that also influences odds of memory retrieval, is spreading activation that has two components, one of which is used here to capture individual differences in attention/working memory capacity, described below.

$$d(i) = e^{A(m,\, t = t(i))} + \alpha \quad (3)$$

Spreading Activation

Spreading activation is the amount of activation that spreads from features of objects in the current attentional focus to those in long-term memory that have matching features and is simply added to the base-level activation B_i (Equation 1, Anderson, 2007). If more spreading activation is fed into a particular chunk, it stands a much higher chance of reaching the threshold for retrieval. There are two parameters in ACT-R that capture this effect that are best explained with an example (W and S_{ji} shown in Equation 1) .

Imagine that a participant is shown a Red Square labeled "X". In this instance, all items in the participant's (or model's) long-term memory that have the features "X", "Red", or "Square" receive spreading activation (Anderson, 2007). There are two important pieces of information in this example that influence how much spreading activation is available. The first is the *number of features*, of which there are three in the example above (label, color, and shape), that have to share the limited amount of spreading-activation. This is denoted by w in ACT-R (Equation 1) and used by Daily et al. (2001) in their model of the Modified Digit Span Task

(MODS), to identify individual differences in working memory capacity. So, the more features there are, the less amount of activation each individual chunk will receive, much like an attentional resource limitation.

The second piece of information, not explicit in the example above, *is the number of memory chunks* in the participant's long-term memory that have *at least one* of those features (label: "X", color: "Red", shape: "Square"). As above, the more memory chunks there are that contain any one or more of those features, the less activation each chunk receives. Given that this is a limited capacity resource (like working memory), each chunk has a reduced likelihood of reaching the retrieval threshold. One can imagine, however, in the case of high resource individuals, the large number of LTM chunks might receive sufficient spreading activation to reach the retrieval threshold, which enables us to model individual differences in capacity. This limited resource is denoted by the parameter S_{ji} (simply S in this text) in ACT-R. The amount of spreading-activation contributed is computed by taking the natural logarithm of the number of chunks in LTM associated with that feature and subtracting it from the constant S (Equation 4). The number of associated LTM chunks is the ratio of the number of memory slots in all chunks ($slots_j$) in LTM that contain that feature to the number of slots ($slotsof_{ji}$)in attended (i.e., currently focused on) chunk (Equation 4, widely known as the fan-effect, Anderson & Reder, 1999). One is added to include the chunk's link to itself. S is our preferred method, of the two described, to capture individual differences in working memory capacity for the RLWM task and the MODS task (our external task for working memory span measure). The reasoning for this choice is described below, in the next section.

$$S_{ji} = S - ln\left(\frac{1 + slots_j}{slotsof_{ji}}\right) \quad (4)$$

Modeling Approach

The goals of the current study are to: 1) test the reliability of RoF and *S* parameters estimates by comparing across the three tasks2) test if RoF and *S* parameters predict strategy selection in the RLWM task, and 3) relate RoF and *S* parameters to learning outcomes. To those ends, we aimed to improve parameter estimation by addressing modeling limitations in our previous studies, identifying a straightforward way of relating parameter values to strategy selection, and learning outcomes. Lastly, a procedure that supports the first aim, externally measures the above described LTM parameters to improve and validate individual parameter values.

Using the Integrated RL-LTM Model of the RLWM Task

The first step in our current approach was streamlining how parameter values might be shown to relate to learning strategies. This can be best achieved by fitting a single multi-mechanism model to all participants and measuring how much RL vs LTM they likely used to learn the RLWM task. One way of achieving this is to fit all participants to a single model that integrates both mechanisms and supplies an unambiguous measure of how much RL vs LTM was used during learning. We have two such integrated RL-LTM models: Biased model and the Meta-RL model. Recall that the Biased model from Chapters 2 and 3 has a parameter (Bias, β) hence its name, that explicitly states what percentage of trials are learned with RL and LTM. The Meta-RL model does not have a bias parameter but learns to use either RL or LTM sub-system by using an RL based meta-learner that chooses which mechanism to use based on their relative success in previous trials. Both of these integrated models have a fair measure of RL-LTM bias, but we took a formal approach to decide between them. This approach even included a comparison to the LTM only model to be well-informed in our decision.

In this comparison process we fit each of the three models (Biased, Meta-RL and LTM only) to participants separately and compared their root-mean squared errors. Separately, here means that the models did not compete with each other. We found that the Biased model fit participants better than the other two. We suspect that when the Biased model is not being penalized for its larger number of parameters by our BIC model selection process, it might explain behavior better. The LTM model was most likely deemed the best model in previous chapters since it had the smallest number of parameters while having relatively low errors. Fitting all participants to a single model does not present a major disadvantage as the integrated model will allow us to capture both "mostly-RL" participants (high percentage of trials use RL) and "mostly-LTM" participants (low percentage of trials use RL) along with those that fall in between.

Improving parameter estimation from RLWM Task

As stated above, we seek to estimate individual reliable parameter values – but we need to address enduring limitations in our modeling efforts. Firstly, we have used a sparse or coarse sampling of the parameter space for simulations which limits our ability to discriminate individuals whose differences might occur at a much finer scale. Secondly, simpler models with fewer numbers of parameters are preferred (hence the necessity of model fit comparison methods that penalize higher numbers of parameters) and Biased RLWM model has six parameters, which needs to be addressed. Thirdly, our parameter recovery (see Chapter 3) procedure also indicated that most LTM parameters were poorly recovered. We reasoned that this may be due to the chance that multiple combinations of parameter values might result in the same curves, with slight variations in fit quality and reduces the level of attributable parameter effects on learning behavior.

We addressed these limitations by reducing the number of Biased Model parameters from six to four by fixing two of them. This reduction in the number of parameters requires much fewer numbers of simulations, which allows for expanded range and dense sampling of the declarative memory parameters. For instance, we were able to increase the number of available values by about a factor of 10, from five values to over fifty. The rest of the improvements in this regard require updates to the Biased Model, specifically to its LTM sub-system, described later in this section.

Independent Tasks and Associated Models

We used two tasks that have been previously shown to estimate model parameters that correlate strongly with task performance and medical assessments of cognitive impairment and have cross-task validation. More importantly they are both based on the ACT-R cognitive architecture which will allow us to "plug-and-play" directly into our RLWM models. The two tasks and their models are described in detail below.

Adaptive Fact Learning (AFL) Task. The AFL task, powered by MemoryLab LLC (www.memorylab.nl) is an adaptive, declarative, paired-association learning task which uses a model-based prediction of the latency and likelihood of paired-association memory retrieval to decide when cues should be presented (Van Rijn et al. 2007, Sense et al. 2018). The goal here is to present cues for a specific association before it is forgotten, but sufficiently spaced during the learning session to take advantage of the spacing effect in declarative memory. The system is ACT-R-based and uses a model developed by Pavlik and Anderson (2008) to make those predictions and decisions. In this model, there is only one free parameter, the rate of forgetting (RoF), which has been shown to be a stable measure of individual memory decay rate and is predictive of learning success (Sense et. al, 2021). The RoF determines how quickly the odds of memory retrieval decrease across time according to the power-law of decay. In ACT-R, it makes up the base-level activation portion of the declarative memory model. In its implementation, the model assumes a common level of retrieval threshold and spreading activation, so these parameters are fixed among participants. Even so, the system has been shown to not only reliably measure individual levels of forgetting , but also distinguishes, with very high accuracy, individuals that have memory related mild cognitive impairment from age-matched healthy controls (Hake & Stocco., 2023).

In a single session of the task, the learning system assumes a default RoF 0.3 and adaptively adjusts this value as participants learn and it gains evidence. If participants respond inaccurately and if they respond much more slowly than predicted, the RoF is increased and predictions for latency and accuracy are adjusted accordingly. If, on the other hand, participants respond accurately and faster than predicted, the RoF is lowered. We plan to use two different categories of arbitrarily associated facts to estimate RoF. Estimates of rate of forgetting vary slightly between tasks (Sense et al., 2016) so two categories that have relatively arbitrary associations between stimulus and response were chosen with the goal of obtaining more stable parameter estimates (described below in materials and methods.

Modified Digit Span Task and Model. The Modified digit span task (MODS) , like many working-memory-span tasks, simply involves presenting digits (from 0 to 9) at regular intervals, with distractors interspersed (letters, in this case), and requests that participants report the digits in the correct order. This version, adapted from Daily et al. 2001, also requires that participants read aloud the target digits and distractor letters to take advantage of articulatory suppression and prevent participants from using mnemonic devices to remember the digits. The MODS ACT-R model we implemented, adapted also from Daily et al., 2001, relies both on the base level activation computation and spreading-activation components of the memory activation equation, described above. However, it relies only on spreading activation to capture individual differences in span size and all other parameters are left fixed at their default ACT-R values or are optimized for the task. The parameter we implement called *S* in ACT-R represents the upper limit for the fan effect which we find is appropriate for capturing individual memory capacity limitations (see description above).

Briefly, we take advantage of this effect to model the strain that larger span trials might project on working memory capacities. That is, if an individual has a large working memory capacity, there will still be a lot of available spreading activation after we subtract the fan effect, where as an individual with a smaller working memory capacity, represented by a smaller value of *S,* will have little to no spreading activation to aid in memory recall, especially at higher span sizes. Here, performance on the task will only be as successful as how well the participant (or model) is able to rehearse the digits and relies only on base-level activation and associated decay rate.

This leads us to the second important feature of the MODS ACT-R model - rehearsal. Daily et al., (2001) asserted that participants rehearse the digits at every chance they get while performing the task (e.g., in-between item presentations). Therefore, their model (and ours) implements a 'respectful but greedy' mechanism for rehearsal of seen target digits. It defaults to rehearsing the digits, in the order that they appeared, as long as there isn't a currently presented stimulus to read aloud and encode. Like the AFL task model above, this model also assumes common group level values for other parameters such as retrieval threshold, memory decay, retrieval noise etc. Refer to the results section below to see how well this model performed in model fits

Updates to the Biased RLWM Model. The Biased model was selected to capture behavior and estimate parameters for subjects because it integrates all the parameters, we would like to measure in one model. However, two major changes need to be implemented: We decided to update the model to implement the Pavlik and Anderson (2008) model modifications that include the spacing effect and directly estimate RoF for comparison with our independent tasks.

Once we integrated the spacing-effects modifications, it was evident that the declarative memory portion of the model was displaying ceiling effects even at high RoF values which was solved by preventing a default, automatic and superfluous behavior in the ACT-R Declarative module that harvests all information put into to other buffers such as the imaginal module. This caused an unrealistic inflation of the base-level activation for chunks resulting in the ceiling-effects.

We observed that our previous version of the model was not also sufficiently capturing memory pressures that result from the increase in set-size. So, inspired by our work on the MODS model, we implemented a minor change to how stimuli are presented to the model that made it more sensitive to increase in set-size.

In the RLWM task, assuming a naive participant (or model in this case), the participant knows what set-size block they are in (3 or 6), and what the stimulus is at the time of stimulus presentation (two features). The difference in set-size was designed to increase pressure on working memory resources. But since the number of features does not change between the two blocks, the increase in memory pressure for set-size 6, compared to set-size 3 cannot be captured by *w,* which is just divided by the number of features. But if we consider *S,* there are double the number of chunks in long-term memory that are associated with block-size 6 compared to set-size 3 that have to share in that limited, individually variable resource, *S*. For this reason, we use *S* for both the MODS task and the RLWM task to be a measure of working memory capacity.

Lastly, as mentioned above, we removed the two noise parameters, LTM and RL, both to reduce the number of free parameters, and therefore make recovery more accurate, and to sample a wider and denser set of parameters.

If we can reliably estimate these parameters (RoF and *S*), we hypothesize that we can get a measure of how much participants are likely to rely on RL vs LTM when they encounter a learning task. According to this hypothesis, what likely happens is that individuals might have learned to use their available memory capacities to their advantage - for example, high decay rate individuals might have learned to engage more procedurally with tasks. This line of thinking resulted from the behavior of one of our integrated models, the Meta-RL model that selected either RL or LTM, dynamically, to engage with the RLWM task depending on the relative success of each of those mechanisms. The meta-learner was more likely to select, for instance, the RL mechanism, if LTM had limited success because of high decay rate, or low spreading activation. Going forward, the integrated Biased RL-LTM model will be referred to as the RLWM model, for simplicity.

Planned Analyses

Recall that the goals of the current study are to estimate individual level model parameters that represent cognitive characteristics and relate them to learning strategy and behavior. To address those goals, we performed three sets of analyses.

First, we aimed to test the stability of individual parameter estimates by extracting the values of RoF and *S* from best fitting models of our three tasks and correlating them across tasks. Given the precedent that RoF and *S,* in their respective tasks, captured individual differences well, we expect that these parameters, extracted from the AFL and MODS tasks respectively, should be correlated with those from the Biased model.

However, based on our previous difficulties with parameter recovery, we also plan to take a second step to see if we can improve RLWM parameter estimation by fitting RoF and *S* using the external tasks. Specifically, this step will use parameter values obtained from the external tasks and follow this procedure: in an attempt to improve *S* estimates, "plug-in" the RoF estimates from AFL task into the Biased model, perform model fits again, and test if *S* estimates now correlate better between Biased model and the MODS task. Next, perform the opposite to improve RoF values, — plug-in *S* values from the MODS model, for each participant, and see if the new RoF values correlate better with those obtained from the AFL task.

Second, to demonstrate that derived parameters are related to learning behavior, and are reliable across tasks, we plan to correlate parameter estimates from all tasks with behavioral data. To further demonstrate that these parameters are reliable, we plan to perform correlations both within and across tasks, using the appropriate parameters.

For the RLWM task, accuracy scores for both the learning and testing phases in the two set-sizes (3 and 6) will be used. For the MODS task, we will use a task accuracy score that is averaged across all span sizes (partial credit unit scoring, PCU, explained below). The only exception here is the AFL task data, as the dependent variable is the parameter estimate, and there are no behavioral results, as such, to use.

In this analysis we expect, RoF estimates from both AFL and Biased model should be negatively correlated with accuracy scores on the RLWM task but that correlations should be slightly stronger with Biased model estimates. The *S* parameter from the Biased model should be strongly positively correlated with accuracy scores in the RLWM task and the PCU score in the MODS task and vice versa. We expect that correlations with-in native tasks to be higher than cross-task correlations .

Third, in addition to explaining behavior, do parameters explain what strategies or combinations of learning mechanisms individual participants land on? This is perhaps too much to ask of our current experimental set up, but we intend to use the bias parameter (β) to create two groups, mostly LTM (LTM group) and mostly RL (RL group) by using 60% as an inflection point and compare means of parameter estimates. This grouping is necessary because β is not a continuous variable and can be interpreted as a categorical variable with eight levels. If we find a difference in the means between the two groups, it will indicate that a specific strategy was used, and that the strategy arose perhaps due to high or low values in parameters.

Materials and Methods

Participants

101 undergraduate students from the University of Washington participated in this experiment. Of those, 90 participants were retained for analyses due to occasional occurrences in which one or more tasks failed during data collection and resulted in incomplete data. Participants were 43% early bilingual, 11% late bilingual and 36% monolingual English speakers recruited through the UW Psychology subject pool (49 females, 4 others, aged 18-35 years, median age 19). Participants received extra credit for their participation. Data were collected after receiving informed consent in one 1.5-hour in-person data collection session.

Behavioral Tasks

Reinforcement Learning Working Memory task

This task (RLWM, Collins, 2018) involves learning stimulus-response associations through a series of 14 blocks. Participants are instructed to respond with a keypress of either ‘C’, ‘V’ or ‘B’ to the displayed images. In 8 of the 14 blocks, participants learn to associate keypresses with three unique images, presented 12 times in random order. In the remaining 6 of the 14 blocks, participants learn to associate 6 unique images each presented 12 times within the block with the key presses stated above. The stimulus-response associations are deterministic, and participants learn through reward (+1 point for correct responses and 0 points for incorrect responses). The same sequence of set-size 3 and 6 blocks was presented to all participants. Following this learning phase, a 10-minute distractor task is administered before a surprise 206-trial test block. Participants make responses without feedback to items taken from both 3- and 6-set learning blocks. Stimulus presentations and data collection were done in MATLAB (mathworks.com) and Psychophysics Toolbox (Brainard, 1997).

Modified Digit Span Task

The modified digit span task (MODS) is a working memory span task adapted from Daily et al., (2001) and Lovett et al. (2000). Like most modified span tasks, it involves intermediary distracting stimuli (distractor letters) in between presentations of the to-be-remembered digits (targets). In this task participants are asked to read aloud all stimuli, both distractor letters and target digits which presents an articulatory suppression and discourages the use of individual mnemonic devices. This version of the task is shorter, lasting about 12 minutes, than the one used by Lovett et al. (2000) or Daily et al., (2001), which lasted an hour. In the current version there is only one condition, span length, which varies from 3 to 6 digits. The original task had an additional condition of distractor length with two levels, hard, at 5 or 6 distractors and easy, at 2 or 3 distractor letters in between the target digits.

Our version used the easy level, therefore only 2 or 3 distractor letters, determined randomly for each span and participant, were used. Each span, 3, 4, 5, and 6, was repeated 8 times for a total of 32 trials. A trial contains a string, which is made up of target and distractor stimuli. An example string of span 3 looks like this, 4-W-H-8-S-J-F-5 (see Figure 1). A trial ends when all the target digits have been presented and input boxes for participant response are displayed. Targets are on screen for 910 milliseconds and distractors are on-screen for 810 milliseconds. Participants are instructed to report just the digits by entering them into the shown boxes (the same number of boxes as the span length in the trial are displayed), in the order they saw them. Participants cannot change answers but reporting answers is self-paced and the next trial is started by the participant when they are ready to continue. All-or-nothing accuracy scores for each span size was computed for model fitting purposes, and a partial unit score (PCU), was used as an individual difference measure of working memory capacity. The PCU is obtained by granting partial credit for all correctly recalled digits as a proportion of the span length (e.g., 3 correct digits out of 6 will be given a score of 0.5, as in Conway et al., 2005). These accuracy scores are then averaged across span sizes and trials, to obtain a single accuracy score.

Figure 1: Schematic of the Modified Span (MODS) Task

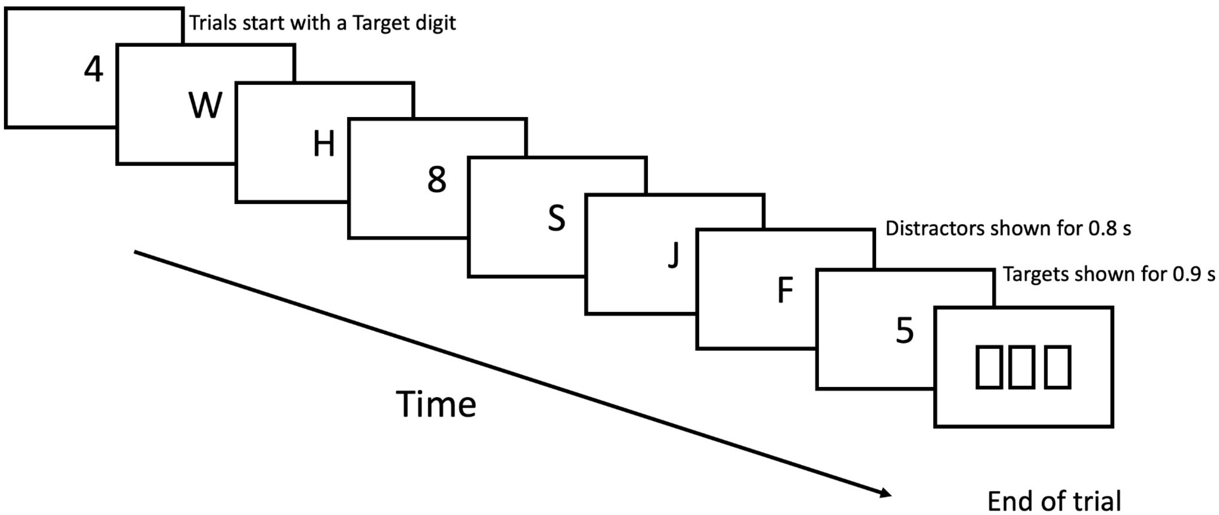


Figure 1:The MODS task takes around 12 minutes. Each trial consists of target numbers and distractor letters. Trials always start with the first target digit and end when input boxes are shown. Targets are shown for 100ms longer than distractors. Transitions between trials are participant paced. There are 32 trials total, consisting of eight repetitions for each of the four span sizes, 3,4,5 and 6.

Adaptive Fact Learning (AFL) Task

The web app (www.memorylab.nl) version of this adaptive fact-learning system (Sense et al., 2018; Sense et al., 2021, Van Rijn et al., 2009) was used to teach participants arbitrary associations in two categories. Here, subjects were presented with an image of types of Candy (first category), or an ancient map of a European city (second category) and were asked to report the associated US state and city name, respectively (see Figure 2 for an example of lesson material). Participants in this task learn through feedback what the correct association is, as there is no study period. As the participants engage with the system and learn, the model refines its estimates of the participants' rate of forgetting, which we obtain as a dependent variable from this task (see Introduction above for details). Additionally, which category, Maps or Candy was administered first was counterbalanced across individuals.

Figure 2: Example stimuli and associated responses for the two-categories in AFL task

Figure 2:Two categories Candy and Maps were used in the Adaptive Fact Learning (AFL) task. The Candy category contained arbitrary associations with US states. Learning takes 8 minutes for each category and was administered at the beginning and end of the experimental sessions. The order was determined randomly for each participant but counterbalanced across participants.

Procedures

In this 1.5-hour study, participants took the three separate tasks in a particular order: AFL task (8 minutes, Maps or Candy), RLWM learning phase (35-40 minutes), MODS task (12 minutes), RLWM testing phase (10 minutes), and a final AFL task (Maps or Candy depending on counterbalance). The 12-minute MODS task serves as a distracting break for the RLWM task (see task description above) but the data is also analyzed. Participants are offered a 1-minute break at each transition point, but they are forced to take the 1-minute break right after the long learning phase of the RLWM task. Note that all sessions begin with obtaining consent and demographics information. The version of AFL administered first, Maps or Candy, is assigned according to a counterbalance for each participant.

Results

Participants performed well on all the tasks overall. As a result, all complete data sets were used for these analyses. However, 11 participants were dropped from cross-task analysis because they had incomplete data in one or more tasks. As described above, we fit the RLWM Biased (hereafter just RLWM) and MODS models to participant data and proceeded with the outlined analyses: 1) testing the reliability of LTM parameters RoF (rate of forgetting) and *S* (spreading-activation) by correlating across tasks (RLWM vs MODS and RLWM vs AFL); 2) testing the predictive power of parameters on behavioral outcomes, and 3) testing the relationship between parameter estimates and best fit strategy as measured by parameter β (bias) obtained from the RLWM model. The results section will follow this organization, but an analysis of behavioral data and model fit quality is described first.

Behavioral Data and Model Fit Quality

RLWM Data Analysis

All participants learned well on the RLWM task, in both the learn and test phases, obtaining well above chance accuracy scores both in the set-size 3 (learn: $M = 0.94$, $SEM = 0.005$; test: $M = 0.73$, $SEM = 0.017$) and set-size 6 conditions (learn: $M = 0.89$, $SEM = 0.009$; test: $M = 0.81$, $SEM = 0.014$) so no participants were dropped for lack of engagement in the task. We followed the same procedures as in Chapters 1 and 2 to fit participants to models, minimizing BIC for model selection but here, as discussed above, we fit only the Biased model. That simply means selecting a combination of parameter values that best explain learning and test behavior. Model fits were remarkably good, as evidenced by a low root-mean-squared-error

(*M:0.064*, *SEM=0.0012*). See Figure 3 for a group-level fit of the learning curves and change from learning to test.

Figure 3: Group level behavioral - model fit curves

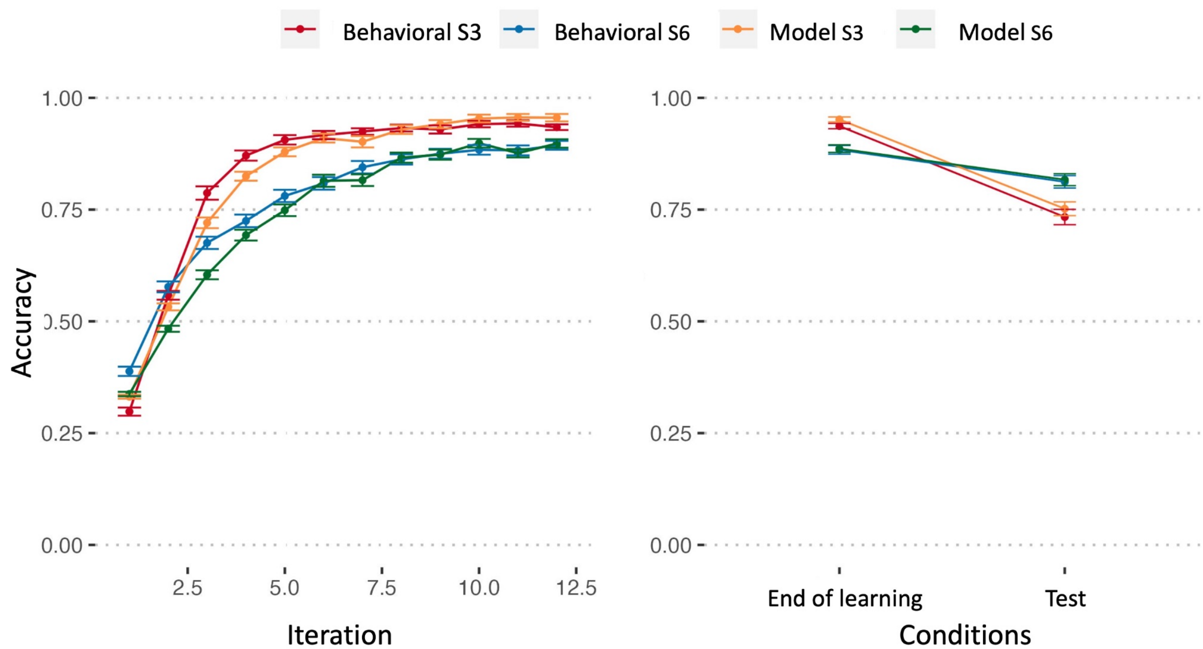


Figure 3: This plot shows the group level model fit curves for learning (left panel) across stimulus iteration (x-axis) and change from end of learning to test (Right panel). Warm colors represent set-size 3 (S3) and cool colors represent set-size 6 (S6).

MODS Data Analysis

Participants performed as expected, even on this abbreviated version of the MODS task. We computed span accuracy, where scoring follows an all-or-nothing method of assigning credit for each span size (sizes 3 to 6), and partial-unit credit scoring (PCU) for all of our participants. Participants (*n* = 97) on average scored 91% for span 3, 78% for span 4, 57% for span 5 and 37% for span 6 (see solid behavioral data curve in figure 4). Their mean PCU score was 83% (*SEM* = 0.012). These span accuracy scores were used to fit the MODS model to participants, where the best fit was selected by minimizing the mean-squared-error, which were very good (RMSE: *M* =

0.097, *SEM* = 0.0042, Figure4). Spreading-activation (*S)* parameter estimates were compared to PCU scores to test how well they predict behavior, discussed below in the model fitting results.

Figure 4: Group level behavioral - model fit curves for the MODS task

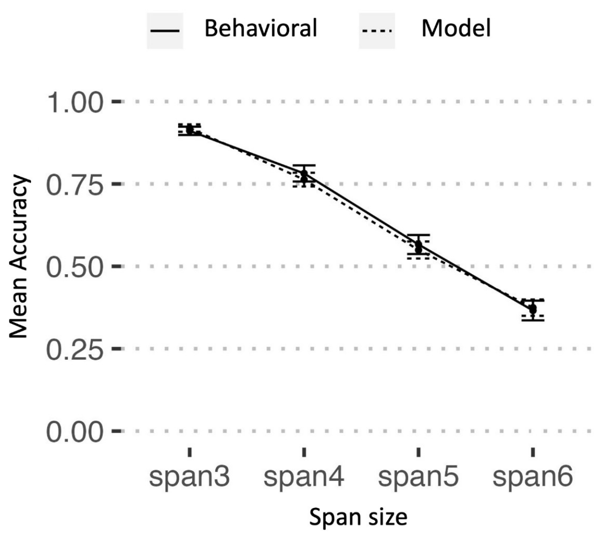


Figure 4: Model fit visualization for the Modified Span Task (MODS). Broken lines are mean accuracy for each span size for participants. Solid line is mean accuracy for each span size for model simulations filtered by best fit to participants.

Adaptive Fact Learning (AFL) Task Data Analysis

Lastly, participants learned the fact associations, on average, in four repetitions, for both the Maps and Candy categories. Accuracy scores were higher for the Candy category compared to those of the Maps category(Maps: *M* = 0.64 *SEM* = 0.025; Candy: *M* = 0.69, *SEM* = 0.023). A separate RoF value was estimated for each category, which was averaged and used in the parameter comparison analysis below. Participants on average had RoF values that were lower for Candy (*M* = 0.29) than Maps (*M* = 0.31), which suggests that the Candy category was easier, or tended to decay slower in participants' memories. The difference between these means is significant ($t(180)$ = , p = 0.00037). Additionally, there was a mildly significant ($\alpha = 0.01$) main effect of Participants' language background (Early Bilingual vs Late Bilingual vs Monolingual) on RoF estimates ($F(2,181) = 4.01$, $p = 0.019$). Recall that the spacing-effects model is embedded in the learning environment and estimates of RoF were made instantaneously as the participants interact with the software and so external model fits were not performed.

Part 1:External Validity of RoF and *S* Parameters

Cross-task parameter comparisons

As stated in the previous section, the analyses here first seek to find if the parameters from the three tasks, RLWM, MODS and AFL are related to one another. To do so, we correlated RoF from RLWM (RoF_{rlwm}) to that from AFL (RoF_{afl}) and *S* from RLWM (S_{rlwm}) to that from MODS task (S_{mods}). Here we found that these cross-tasks parameters were not correlated at all (*S* parameters: $r = 0.04$; RoF parameters: $r = -0.04$, see peach colored lines in figures C1 and C2, respectively). This suggests that parameter values do not reflect cognitive characteristics that are stable over different contexts.

However, given that the RLWM model has multiple free parameters, we planned to improve cross-task stability of parameters by using the external estimates to reduce the parameter space for the RLWM model. That is, to improve stability for the *S* parameter, we fixed RoF values in the RLWM model using those estimated from AFL, and fit models to participants again. The same procedure was performed for the RoF parameter, by fixing *S* values in RLWM models using those estimated from the MODS model.

Figure 5: Cross task correlations of *S* and RoF parameters.

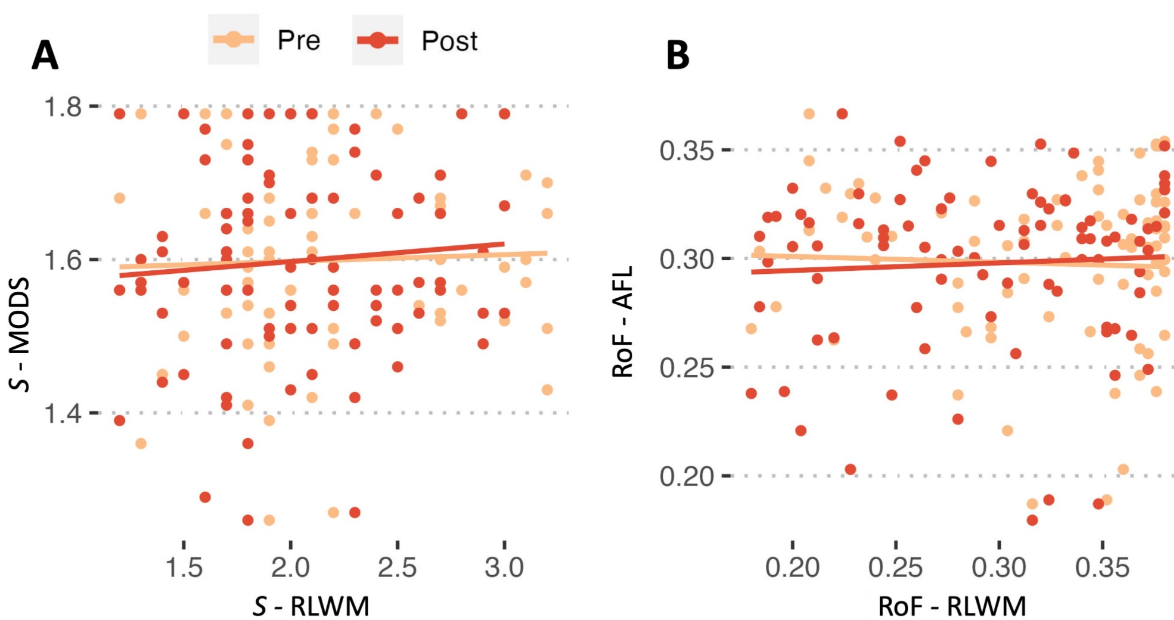


Figure 5: shows cross-task correlations for the MODS task (5A) and its spreading activation parameter *S* and correlations (5B) for the Adaptive Fact Learning task (AFL) for its parameter RoF (rate of forgetting) with those estimated from the RLWM task.

Fixing RoF and comparing *S* cross-task.

Since the RoF parameter space in the RLWM simulations was not sampled with as high a degree of precision as that from AFL, the closest RoF_{rlwm} value was determined for each participant's RoF_{afl}, to three decimal places, using the existing set of simulations. Once RoF was fixed, the reduced number of RLWM models were fit to those participants to see if estimates of *S* correlated better with those of the MODS task. As shown in Figure 5A(red trend line), there was negligible to no improvement in correlation between the two estimates (new $r = 0.09$). In fact, fixing RoF in RLWM resulted in overall poor model fits.

Fixing *S* and comparing RoF cross-task.

For the same reason stated above, the closest simulated S_{rlwm} to S_{mods} was obtained for the purpose of fixing that variable before performing fits again. An additional step was necessary here: both sets of *S,* RLWM and MODS, needed to be scaled before finding the closest match since the range of parameters were optimized for each task and therefore fell outside of bounds for each other. Similar to the RoF results above, correlation between RoF_{rlwm} and RoF_{afl} were poor to none (new $r = 0.06$, Figure 5B, red trend line) after fixing *S* parameter values for each individual.

These results are a departure from what was expected given the results Daily et al., Lovett et al., where robust cross task effects were observed. It should be noted that we had a more ambitious experiment in which we had hoped to estimate *multiple*, individual parameters using a small set of tasks. In Daily et al., (2001) only a single free parameter was estimated but was demonstrated to be highly reliable across two related tasks and their models - a parameter estimated using the MODS task and model allowed for great fit on a model of an n-back task. We compared fits for the RLWM task before and after fixing the two LTM parameters, but it resulted in worse fit.

Part 2: Predicting Behavior from Parameter Values

74 unique sets of parameters fit the 90 participants on the RLWM task, which means some participants overlap in what parameter values were estimated for them. But this is a much higher number of unique parameter sets, suggesting that we have been able to capture more variability with the current set of parameter values. Similarly, 35 unique *S* parameter values fit 90 participants in the MODS task, signifying more overlap than RLWM parameters. *S* and RoF parameters were used to predict behavior within native and across independent tasks.

Spreading Activation (S) Effects

In the RLWM task, S_{rlwm} ranged from 1.2, at which point spreading-activation is zero, even at the smallest fan-effect (see spreading-activation description and equation above), to 3.2, above which would produce ceiling effects (but was also limited to reduce constraints on computation power). The distribution of best-fit S_{rlwm} values has a slight positive skew, peaking at 1.8, with a mean of 2.06. While the parameter from the MODS model, S_{mods}, had a different range, 1.2 to 1.8, which was optimized for the MODS task (models showed ceiling effects at values higher than 1.8), and was negatively skewed with a mean of 1.6. Regardless of this difference in range, we expected that, if we are indeed quantifying stable cognitive traits, a correlation between the estimates and behavior should be evident.

Figure 6: Correlation between MODS model spreading-activation and WM score

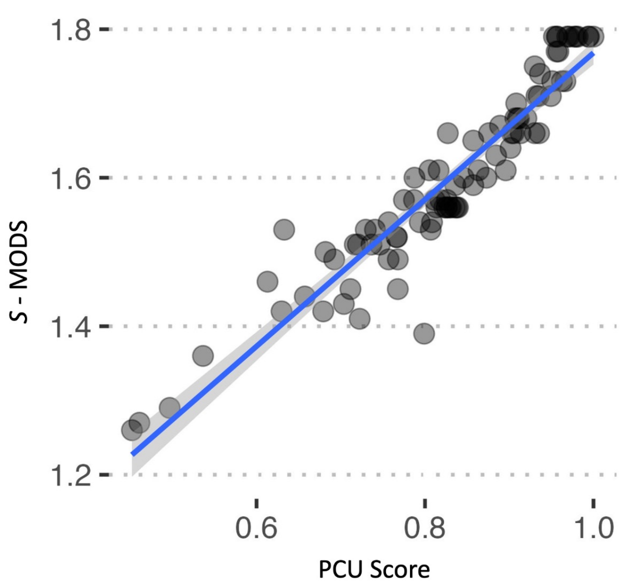


Figure 6: Partial credit unit score (PCU) was used to measure working memory performance on the MODS task. This was strongly correlated with the estimated spreading-activation parameter *S* in the MODS model (r = 0.94).

We correlated *S* values with behavioral outcomes both in the RLWM task (accuracy scores for learning and test in both set-size 3 and 6 conditions) and the MODS task (PCU score). We found that S_{rlwm} was a good predictor of learning on the RLWM task only for set-size 6 (learn: r = 0.29*,* test: r = 0.48) compared to set-size 3 (learn: r = 0.2, test: r = 0.38) and were better for testing accuracy compared to learning accuracy (differences not significant). But S_{rlwm} was not correlated with PCU scores (r = 0.08*)*, violating our expectation that this parameter would be predictive of learning across tasks. Given that, S_{mods} also does not predict accuracy scores on the RLWM task: set-size 6 (learn: r = 0.09*,* test: r = 0.109); set-size 3 (learn: r = 0.06, test: r = 0.07). See confusion matrix in Figure 7. Lastly, S_{mods} remarkably captures individual performance on the MODS task, resulting in a very high correlation with the PCU score (r = 0.94*;* Figure 6).

Figure 7: Table of correlation coefficients between RLWM conditions and parameters

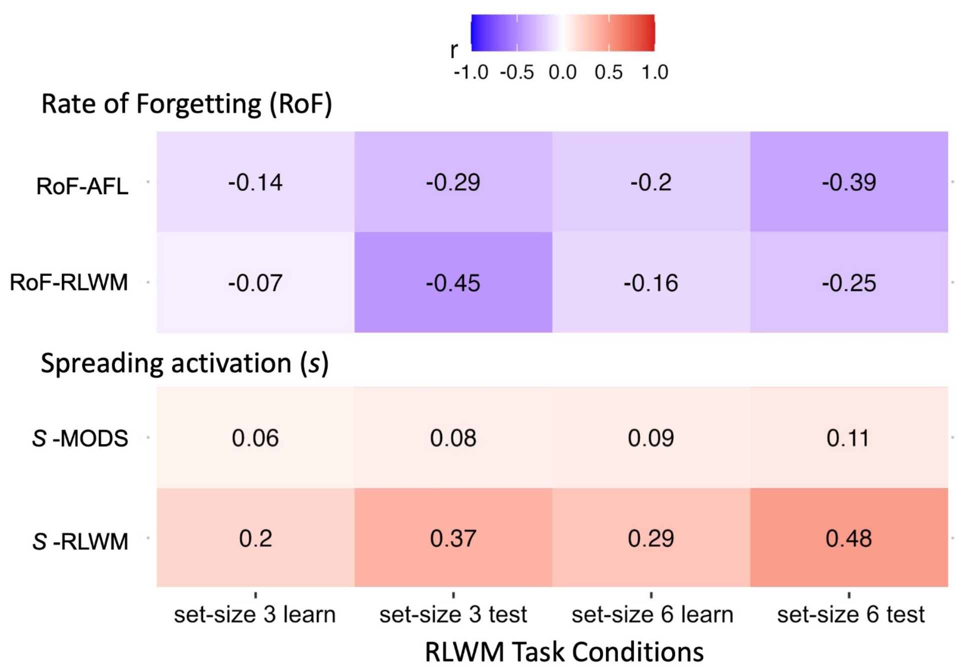


Figure 7: This figure shows correlation coefficients between RLWM task accuracy scores in all conditions (x-axis) with RoF and *S* parameter estimates from Adaptive Fact Learning (AFL) task, MODS task and RLWM task (y-axis). Positive correlations approach red and negative correlations approach blue.

RoF Parameter effects

The rate of forgetting parameter (RoF) has been shown to capture individual level stable memory decay rates in the AFL task. We utilized the same parameter in our RLWM model by implementing the Pavlik and Anderson (2008) spacing-effect model. A range of 0.18 to 0.38 was used in RLWM model simulations to mirror the span of RoF (hereafter RoF_{rlwm}) values observed in healthy, young participants (e.g., Sense et al., 2016). This parameter space is densely sampled in RLWM simulations (in increments of 0.004 in the above specified range) with the goal of approaching the precision of estimates from AFL.

The distribution of fit RoF_{rlwm} values was heavily negatively skewed, where a more normal distribution around 0.30 was expected (M = 0.32), given AFL task data. Estimates from the two AFL (RoF_{ss}) task categories were averaged together to stabilize estimates and resulted in a mean RoF_{afl} of 0.30 across participants. The distribution of RoF_{afl} values was moderately normally distributed, with a negative skew, which replicates results obtained from other studies for this age group.

As above, RoF estimates from the two tasks were correlated with RLWM task accuracy scores. We found that, as expected, RoF values were negatively correlated with accuracy. RoF_{rlwm} values were overall *less* related to RLWM accuracy than RoF_{afl} for all accuracy scores but one. Correlations with RoF_{rlwm} were much higher for test accuracy scores (set-size 3: r = -0.45, set-size 6: r = -0.25) than learning score (set-size 3: r = -0.07, set-size 6: r = -0.20). RoF_{ss} was mildly correlated with set-size 6, especially for test accuracy (learn: r = -0.20, test: r = -0.39), but correlation was higher for set-size 3 (learn: r = -0.14, test: r = -0.29). See correlation table in Figure 7.

Part 3:relating parameter estimates to strategy

RLWM Bias-parameter Analysis

The bias parameter, β, ranges from 10% to 90%, in 10% increments, and represents the percentage of trials that used the RL sub-model to learn associations, in that set of simulations. After fitting models to participants, we found that they were nearly evenly split between less than 30% RL use, meaning mostly LTM (57% of subjects), and more than 70% RL use (41% of subjects, see Figure 8). Even then, a majority of these participants, about 80%, were at the 10 and 90% values. Only 3 participants fit simulations that used RL for 40 to 60% of trials. This is a slight departure from what our previous data have shown, where participants tended to largely fall on the 20 and 40% β values. Note that our earlier set of simulations (see Chapter 1) explored only 20 to 80% in increments of 20% which limited the amount of variability we can capture. This was remedied in this current study, and we took advantage of the polar nature of where participants landed to group participants into 'mostly LTM' (52 participants) and 'mostly RL' (38 participants) using 60% as a cut-off for parameter comparisons and its relationship to strategy.

Figure 8: Distribution of participants' best fit β parameter values.

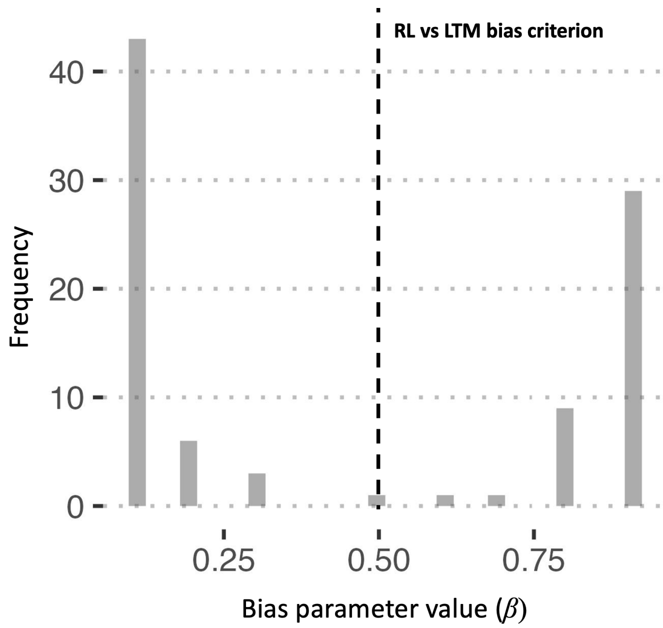


Figure 8: The β (bias) parameter represents the proportion of trials that used RL in the RLWM task. This fit value represents strategy for each participant. This parameter was sampled at 0.1 increments from 0.1 to 0.9 for simulations. The criterion line of 0.5 was used as an inflection point for categorizing participants as 'mostly RL' (higher than 0.5, inclusive) and 'mostly LTM' (lower than 0.5, exclusive).

RL learning rate (α) effects

The only remaining parameter on the RL portion of the integrated model (softMax τ, or RL noise was fixed to reduce the parameter space) spans from 0.05 to 0.25 in 0.05 increments and has remained unchanged since our first modeling attempt in Chapter 1. Since few subjects fit the RL-only model in Chapter 1, this parameter was not selected for optimization as the others. The distribution of fit values of α were positively skewed, with a mean of 0.181, and most participants fit the highest value of 0.25. The parameter is mildly correlated with learning behavior and was more related to set-size 6 (learn: $r = 0.31$; test $r = 0.28$), than set-size 3 (learn: $r = 0.32$; test $r = 0.06$). Notice that correlations are also higher for the learning phase, compared to testing phase, which is contrary to the pattern we have seen with other parameters.

Relating strategy (β) to parameter estimates

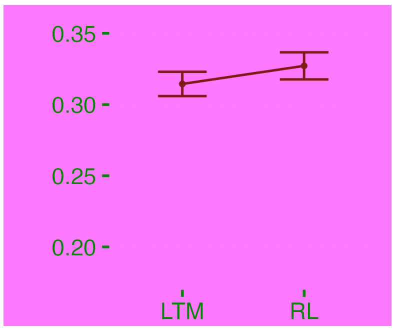
To test if estimated parameters, both those from the RLWM task and the two independent tasks, are related to what proportion of RL or LTM, we took the RL and LTM bias groups and performed a Wilcoxon Rank Sum test, given that parameter distributions violate multiple assumptions for a Student's t-test, comparing the mean parameter values for each parameter (Figure 9). To support our hypothesis that parameter values or cognitive characteristics, influence, strategies two conditions have to be satisfied. We would expect, 1) a statistically significant difference between the groups in parameter values and, 2) mean differences between the groups, for each parameter, should point in the direction that supports a specific strategy. For example, for the Rate of Forgetting parameter, the mean value should be lower for the LTM group compared to the RL group, suggesting that learners are more likely to rely on declarative strategies because of their slower rates of forgetting.

We found that both RoF_{rlwm} (w = 830, p = 0.196) and S_{rlwm} (w = 1086, p = 0.424) parameter values were not significantly different between the two groups. The differences in the means were very small but somewhat in the directions we would expect (see figure 9A and 9B). The externally estimated parameter, RoF_{afl} (w=1340, p = 0.004) was significantly different between the two groups at $\alpha = 0.01$ but S_{mods} (w=1340, p = 0.08) was not. Both had appreciable differences between the two groups albeit in directions we did not expect (e.g., RoF_{afl} values were *higher* for the LTM group, Figure 9D and 9E). Finally, and strikingly, the RL-learning rate parameter was the best and strongest predictor learning strategy (w=308, $p < 0.0001$). While this parameter is the least discussed parameter in this study, it might be the key to understanding the patterns of modeling results we have obtained.

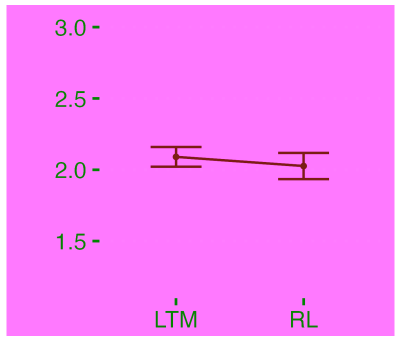
Figure 9: Comparison of fit parameter values by strategy.

Within-task parameters

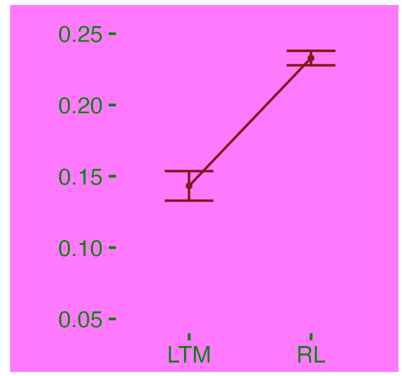
A RoF - RLWM



B *S*-RLWM

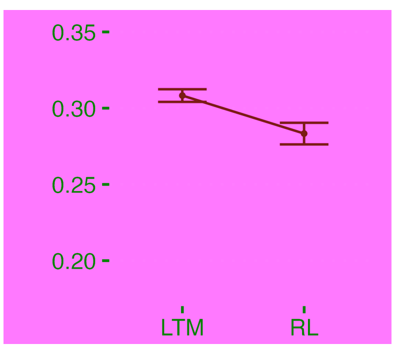


C RL-learning rate



External task parameters

D RoF- AFL



E *S*-MODS

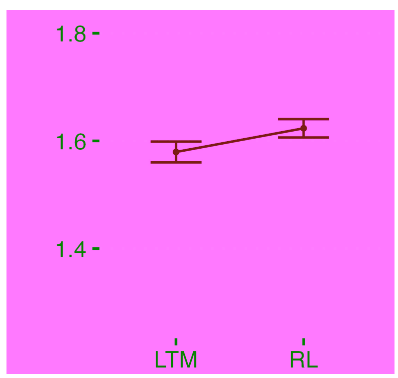


Figure 9: This panel of plots shows the differences in estimated parameter values for the 'mostly LTM' group and the 'mostly RL' group (x-axes). The groups were created from the β parameter on the RLWM task using 0.6 as criterion for categorization. Top panel (A, B and C) are within task parameters (i.e., from the RLWM task). Bottom panel are external task parameters (D: AFL task; E: MODS task). The y-axes show the parameter values. Error bars show standard errors.

Discussion

The current study sought to test the reliability of parameter values estimated from the RLWM task and model by comparing them to those obtained from independent tasks. It was important to do so because our specific motivations were to explore whether the measured parameters reflected stable cognitive characteristics of individual learners, which might drive the learning mechanisms they engaged in the RLWM task. We found that the specific LTM parameters we measured did not correlate across tasks, but that these parameters were moderately successful at explaining learning behavior within task. Surprisingly, externally estimated LTM parameters were better at distinguishing strategy in the RLWM task.

The first of our three goals for this study involved measuring individual parameters which we have accomplished with far greater success than we have in our previous modeling efforts. In this study, models fit behavior better, and the expanded, more densely sampled parameters allowed for capturing more diversity in our learners - more than double the amount we were able to capture in our previous efforts. While successful in these aspects, we learned that parameter estimates are perhaps more meaningful within their specific contexts and had little to no predictive power outside of those contexts. In other words, parameter estimates did not represent more general and stable cognitive characteristics.

The two independent tasks we chose, the AFL task and the MODS task have each been shown to capture vital cognitive characteristics - declarative memory decay rate and working memory capacity, respectively. Memory decay rates, estimated from the Pavlik and Anderson (2008) spacing effects model have been used to teach facts more efficiently and also represent levels of cognitive decline in a population with mild cognitive decline. While we estimated the same parameter on the RLWM task for our participants, we found that estimates did not correlate across the two tasks, and rates of forgetting were meaningful only within the context of the RLWM task, accounting for some of the variability in learning success. The same was true for the spreading activation parameter (*S*) which was used by Lovett et al. (2001) to represent working memory capacity, estimated in a span-task, and had high predictive power in a similar but independent task, the n-back task. Our estimates of *S*, however, were only meaningful within the context of the RLWM task and were not validated as a task independent measure of working memory capacity.

While we conclude that parameter estimates as measures of individual differences are likely yoked to the tasks that produced them, there are perhaps model and task design decisions that might have limited the level of generalizability of these parameters. For instance, our RLWM model estimates multiple parameters at once which spreads the ascribable parameter effects to multiple constructs - memory decay rate, RL learning rate, working memory capacity etc. This means that this model is not best at measuring the contribution of any one of these parameters alone. We tried to lessen this impact by reducing the number of parameters from 6 to 4, by fixing the noise parameters, and again by fixing some parameters using external estimates to improve the others. But this only resulted in worse fits suggesting that the large set of parameters is not the only drawback to our approach. This line of thought was motivated by the fact that the MODS model and the AFL model had only one free parameter.

The results of this paper afford a much wider-angle view of the work we have done so far. Four results from across chapters 2 to 4 point to other model design choices that might have produced this pattern of results. The popularity of the LTM model in Chapter 2 convinced us that declarative memory parameters, perhaps most influenced learning. Secondly, The RL only model and the integrated models, on visual inspection fit the minority of learners they represented exceptionally well. Thirdly, parameter recovery analysis pointed out that LTM parameters are poorly recovered, while RL parameters were recovered well, which motivated a lot of the decisions for chapter 3. And lastly, the exceptionally good fit of the biased model suggests that how the two mechanisms are combined, and what each portion contributes to learning might be the reason for that good fit. These four observations, especially the last one, put together might explain the success and failures we have observed thus far.

The LTM model, while being popular, was limited in explaining learning behavior. For instance, learners were very fast in acquiring the associations, a characteristic Collins (2018) and others might ascribe to engagement with WM, as there was a lot of information lost during the testing phase. But the LTM model never quite achieved this rapid learning even at high levels of spreading-activation, and we believe the LTM was best fit for these subjects because of the amount of change from the learning phase to the testing phase. The RL model on the other hand captured this rapid learning, perhaps owing to the range of learning rate values we used but fit few people in Chapter 1 because an RL only model predicts no decay across a brief delay.

The integrated models beautifully captured these two effects because, we suspect, rapid learning was facilitated by the RL sub-mechanism and decays were evident across the brief delay between learning and test. There is an important difference between the two integrated models, one which might have resulted in our ultimate decision to use the biased model over the Meta-RL model — information sharing between the two sub-models. The Meta-RL model has sub-models that are encapsulated and do not share learned associations, but the Biased model does. The RL portion of the Biased model shares information with the LTM portion, therefore a loss only occurs if they decay due to a time delay or have not received sufficient spreading activation. This information sharing might explain why the Biased model fits well, and how well information is acquired during learning. That is, there are now two robust systems RL (learns quickly) and LTM (RL contributions also generate memory traces) and decay occurs when LTM contributes responses for some of the trials. It is not clear if this is indeed how memory systems interact, even though it shares some aspects of how the basal ganglia affect declarative memory encoding, retrieval, and WM gating (see Scimeca and Badre, 2012; Stocco et al., 2010; Ashby and Crossley, 2010), or if this is one way a multi-mechanism learning system can be built. Regardless, we might have inadvertently created a model where RL and its parameter, α, including the fact that it is highly reliable, is the driving force behind strategy choices, hence its highly predictive power of bias (β) that we have observed in this study, and the diminished effects of the LTM parameters.

In conclusion, it should be stressed that parameters have great utility for capturing individual characteristics, but they are likely affected by model design and are best interpreted in the specific contexts of those models. This might also mean that future goals should aim to measure cognitive characteristics and task demands *together* as these are likely intimately tied, and task free parameters might not be the best predictors of behavior or learning success in a given environment.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford Univ. Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186.

Ashby, G. F., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*(1), 1–12. https://doi.org/10.1016/j.nlm.2010.03.001

Collins, A. G. E. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *Journal of Cognitive Neuroscience*, *30*(10), 1422–1432. https://doi.org/10.1162/jocn_a_01238

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, *25*(3), 315–353. https://doi.org/10.1207/s15516709cog2503_1

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*(1), 284–294. https://doi.org/10.1016/j.cognition.2007.07.001

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research*, *1*(2), 99–118. https://doi.org/10.1016/S1389-0417(99)00012-1

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117. https://doi.org/10.1037/1076-898X.14.2.101

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

Scimeca, J. M., & Badre, D. (2012). Striatal Contributions to Declarative Memory Retrieval. *Neuron*, *75*(3), 380–392. https://doi.org/10.1016/j.neuron.2012.07.014

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

Sense, F., Meijer, R. R., & Van Rijn, H. (2018). Exploration of the Rate of Forgetting as a Domain-Specific Individual Differences Measure. *Frontiers in Education*, *3*, 112. https://doi.org/10.3389/feduc.2018.00112

Sense, F., Van Der Velde, M., & Van Rijn, H. (2021). Predicting University Students' Exam Performance Using a Model-Based Adaptive Fact-Learning System. *Journal of Learning Analytics*, *8*(3), 155–169. https://doi.org/10.18608/jla.2021.6590

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

Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review*, *117*(2), 541–574. https://doi.org/10.1037/a0019077

Stocco, A., Prat, C. S., & Graham, L. K. (2021). Individual Differences in Reward-Based Learning Predict Fluid Reasoning Abilities. *Cognitive Science*, *45*(2), e12941.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

van Rijn, H., van Rijn, D. H., & van Woudenberg, M. (n.d.). *Passing the Test: Improving Learning Gains by Balancing Spacing and Testing Effects*.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Xu, Y., Stocco, A., (2021). Recovering Reliable Idiographic Biological Parameters from Noisy Behavioral Data: The Case of Basal Ganglia Indices in the Probabilistic Selection Task. *Computer Brain & Behavior* 4, 318–334.

Chapter 5: General Discussion

The trio of papers presented here explore fundamental aspects of individual differences in skill learning:1) do learners use different strategies? 2) Are these strategies stable across time and at different task requirements and, 3) can we identify latent cognitive characteristics and relate them to learning success or strategy? We addressed these questions through idiographic computational models and associated parameter estimates, an approach that pushed the boundaries of what has been done in the past with computational models. What we found paints a complicated picture of connections between individual learners, strategies, task requirements, and models.

In the first experiment (Chapter 2), we demonstrated that participants do deploy different memory strategies, and the idiographic modeling technique was instrumental in uncovering that. Few modeling studies compare multiple models fit to individuals, as we have done, which limits what we are able to understand about individual strategies. By adopting the idiographic technique, we found that different strategy models fit subsets of our participants best. Given the task and models we built, we have sufficient evidence that shows that the best fit models were strongly preferred over the alternatives (this includes sets of parameter values within the same model). This pattern continued, and with increased fidelity, when we fit only one multi-strategy model to all participants in the third experiment (Chapter 4), where we observed an almost 60-40 split in strategy preference between LTM and RL, respectively.

In the second experiment (Chapter 3), we went a step further and changed our approach to uncover more information about individuals – how they might respond to different task demands and contexts. In other words, we wanted to examine individual learning dynamics and learner-task interactions, as these affect ultimate learning success. Given that most learners engage in metacognition, which affects strategy, and that the learning task we used was designed to invoke different strategies, we fit the two task conditions separately. Additionally, we split up the learning task into two halves and fit those separately as well. What we found points to an intricate link between learner and task that was difficult to tease apart.

It seems like a majority of participants likely engage the same strategy across the long learning task *and* the two conditions. But some did not. For those who did not use the same strategies, arriving at an LTM strategy led to better performance, albeit in limited conditions, compared to other strategies. This suggests an interaction between task demands and individual characteristics and not always in the direction we predicted.

The RLWM task was designed to invoke RL use for the difficult condition and LTM for the easier condition. We did not observe such a separation, but the easier condition was associated with more diversity in strategy use, where we observed participants falling on a spectrum of LTM only, a combination of LTM and RL, and an RL only group. But the RL group vanished in the case of the more difficult condition and those in the middle skewed more LTM. This aligns with what we know about metacognition and cognitive characteristics. For instance, learners with larger working memory capacity have more capability to compare different strategies and evaluate their performance. Since the easier condition on the RLWM task is more sparing of working memory resources, participants are able to engage in more metacognitive processes and the more difficult situation limits that ability. We were, however, limited by the lack of density of parameter values we were able to estimate for each participant and so could not clearly connect latent cognitive characteristics to strategy and learning outcomes. In the last experiment, we addressed some of these limitations to draw a clearer picture.

Our last study took a robust approach to estimating and validating parameter values. We were able to estimate individual parameters with increased precision on the RLWM task, and estimated LTM parameters independently with external tasks, for validation, as these were poorly recovered on a parameter recovery analysis.

This analysis yet again revealed another level of complexity — model design has a large impact on what values parameters take when fit to participants. In short, parameter values are reliable measures of individual latent cognitive characteristics only within the confines of the models or contexts that produced them. Parameter values were moderately predictive of learning outcomes only within tasks and not across tasks. For instance, spreading activation estimates from the RLWM model predicted learning accuracy in the RLWM task but not in the MODS task (our chosen external working memory capacity task). The opposite was also true, the MODS task spreading activation estimates did not predict RLWM accuracy. In the same vein, most parameters were poorly related to strategy measures in the RLWM task. RL learning rate was the sole significant predictor of strategy even while we obtained superior model fits compared to our initial attempt in Chapter 2. We designed great models that explained behavior very well but inadvertently landed on a design that relied heavily on one parameter and not strongly on the diverse set of parameters, spanning memory mechanisms that we had hoped to capture for all individuals.

In conclusion, individual differences in strategy selection exist in skill learning. These strategies are dynamic in an individual, are related to cognitive characteristics and are affected by task demands. Idiographic computational models are a great tool for uncovering these strategies and linking them to estimated latent variables, but great care must be taken in their design and ultimately, are most useful within the context of task and design assumptions.

General References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Anderson, J. R. (2000). Learning and memory: An integrated approach (2nd ed.). *John Wiley & Sons Inc.*

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford Univ. Press.

Ashby, G. F., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*(1), 1–12. https://doi.org/10.1016/j.nlm.2010.03.001

Baddeley, A. D., & Hitch, G. J. (2007). Working memory: Past, present… and future. *The cognitive neuroscience of working memory*, 1-20.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test.* (p. 27). Psychological Corporation.

Collins, A. G. E. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *Journal of Cognitive Neuroscience*, *30*(10), 1422–1432. https://doi.org/10.1162/jocn_a_01238

Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working Memory Load Strengthens Reward Prediction Errors. *The Journal of Neuroscience*, *37*(16), 4332–4342. https://doi.org/10.1523/JNEUROSCI.2700-16.2017

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*(1), 284–294. https://doi.org/10.1016/j.cognition.2007.07.001

Delgado, M. R., & Dickerson, K. C. (2012). Reward-Related Learning via Multiple Memory Systems. *Biological Psychiatry*, *72*(2), 134–141. https://doi.org/10.1016/j.biopsych.2012.01.023

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*(11), 1611–1618. https://doi.org/10.1038/nn1575

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition:* Defense Technical Information Center. https://doi.org/10.21236/ADA084551

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.

Gluck, M. A. (2002). How do People Solve the "Weather Prediction" Task?: Individual Variability in Strategies for Probabilistic Category Learning. *Learning & Memory*, *9*(6), 408–418. https://doi.org/10.1101/lm.45202

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485. https://doi.org/10.1016/0896-6273(95)90304-6

Hikosaka, O., Yamamoto, S., Yasuda, M., & Kim, H. F. (2013). Why skill matters. *Trends in Cognitive Sciences*, *17*(9), 434–441. https://doi.org/10.1016/j.tics.2013.07.001

Hill, N. M., & Schneider, W. (2006). Brain Changes in the Development of Expertise: Neuroanatomical and Neurophysiological Evidence about Skill-Based Adaptations. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 653–682). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.037

Just, M. A., & Carpenter, P. A. (n.d.). *A capacity theory of comprehension: Individual differences in working memory*.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science*, *4*(6), 533–550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–324. https://doi.org/10.1016/j.tics.2010.05.002

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A Neostriatal Habit Learning System in Humans. *Science*, *273*(5280), 1399–1402. https://doi.org/10.1126/science.273.5280.1399

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

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, *30*(2), 294–301. https://doi.org/10.3758/BF03195290

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research*, *1*(2), 99–118. https://doi.org/10.1016/S1389-0417(99)00012-1

McDonough, I. M., Enam, T., Kraemer, K. R., Eakin, D. K., & Kim, M. (2021). Is there more to metamemory? An argument for two specialized monitoring abilities. *Psychonomic Bulletin & Review*, *28*(5), 1657–1667. https://doi.org/10.3758/s13423-021-01930-z

Miyake, A., & Friedman, N. P. (1998). Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retentio* (pp. 339–364). Lawrence Erlbaum Associates.

Newell, A., & Rosenbloom, P. S. (1980). Mechanisms of Skill Acquisition and the Law of Practice.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

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*. 7.

Prat, C. S., Yamasaki, B. L., & Peterson, E. R. (2019). Individual differences in resting-state brain rhythms uniquely predict second language learning rate and willingness to communicate in adults. *Journal of Cognitive Neuroscience*, *31*(1), 78-94.

Prat, C. S., Madhyastha, T. M., Mottarella, M. J., & Kuo, C.-H. (2020). Relating Natural Language Aptitude to Individual Differences in Learning Programming Languages. *Scientific Reports*, *10*(1), 3817. https://doi.org/10.1038/s41598-020-60661-8

Prat, C. S., Seo, R., & Yamasaki, B. (2015). The Role of Individual Differences in Working Memory Capacity on Reading Comprehension Ability. In *Handbook of Individual Differences in Reading*. Routledge.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Non-learners during Reward-Based

Decision Making. *The Journal of Neuroscience*, *27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schultz, W., Dayan, P., & Montague, P. R. (n.d.). *A Neural Substrate of Prediction and Reward*. 8.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction 1st edition.

Sense, F., Van Der Velde, M., & Van Rijn, H. (2021). Predicting University Students' Exam Performance Using a Model-Based Adaptive Fact-Learning System. *Journal of Learning Analytics*, *8*(3), 155–169. https://doi.org/10.18608/jla.2021.6590

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Shute, V. J. (1991). Who is Likely to Acquire Programming Skills? *Journal of Educational Computing Research*, *7*(1), 1–24. https://doi.org/10.2190/VQJD-T1YD-5WVB-RYPJ

Sloman, S. A. (n.d.). *The Empirical Case for Two Systems of Reasoning*.

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

Tenison, C., Fincham, J. M., & Anderson, J. R. (2016). Phases of learning: How skill acquisition impacts cognitive processing. *Cognitive Psychology*, *87*, 1–28. https://doi.org/10.1016/j.cogpsych.2016.03.001

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *30*(6), 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

van Rijn, H., van Rijn, D. H., & van Woudenberg, M. (n.d.). *Passing the Test: Improving Learning Gains by Balancing Spacing and Testing Effects*.

Yamasaki, B. L., Stocco, A., & Prat, C. S. (2018). Relating individual differences in bilingual language experiences to executive attention. *Language, Cognition and Neuroscience*, *33*(9), 1128–1151. https://doi.org/10.1080/23273798.2018.1448092