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

Theodros M. Haile

Department of Psychology, University of washington Seattle

theodros@uw.edu

Chantel S. Prat

Department of Psychology, University of Washington Seattle

csprat@uw.edu

Andrea Stocco

Department of Psychology, University of Washington Seattle

stocco@uw.edu

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

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-cogntive 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) tast (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).

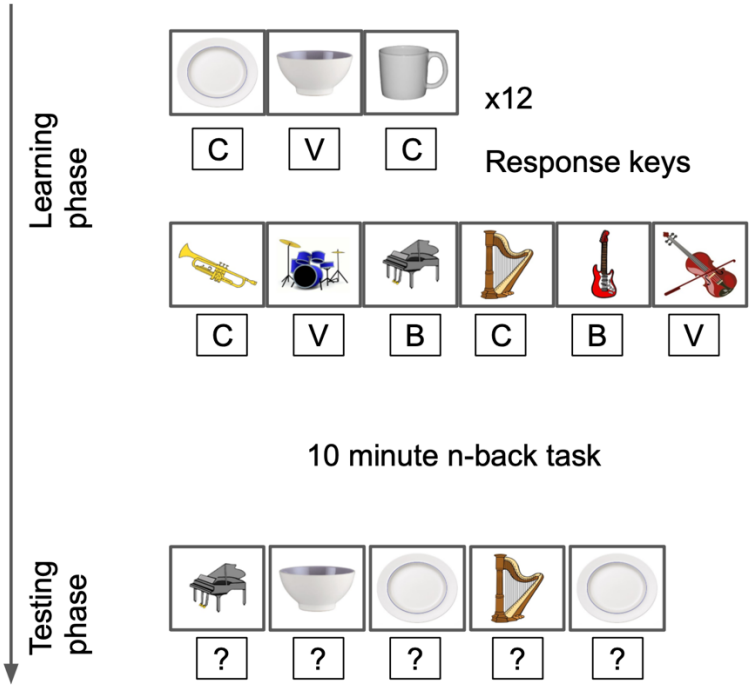

Figure 1: Schematic of the RLWM task. The images are examples of actual stimuli used in the task.

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 half the blocks participants learn to associate keypresses with three unique images, presented 12 times in random order, and in the other half, 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 deterministic, and participants learn through trial-and-error and with explicit feedback (+1 point for correct responses and 0 points for incorrect responses). 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 the 3- and 6-set learning blocks. 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,

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

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.

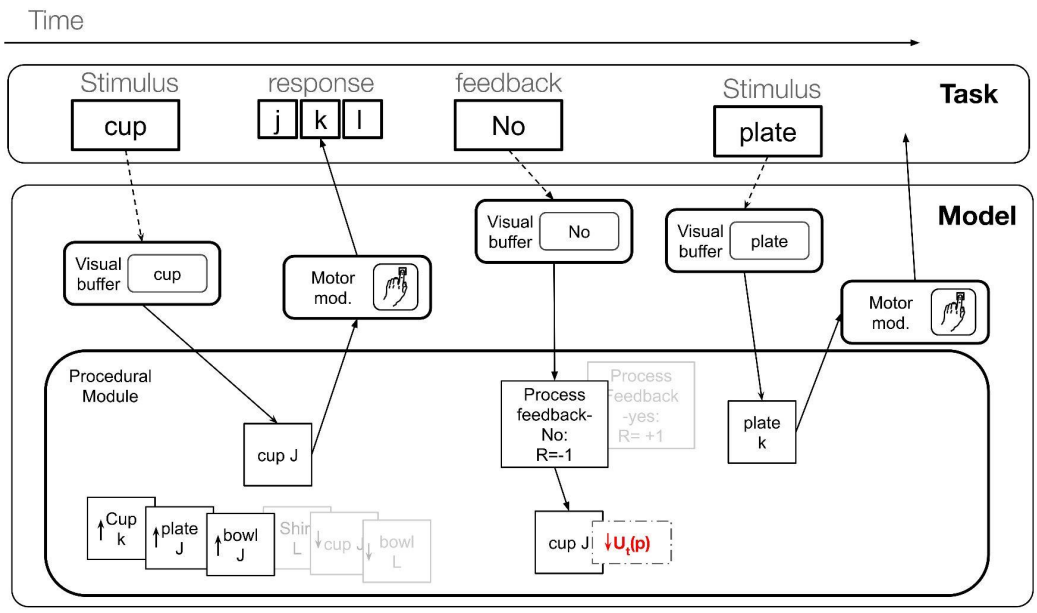

Figure 2: Overview of the procedural RL model as implemented in ACT-R. The feedback, in the example case, 'No', is sent to the RL-based utility learning system in ACT-R which reduces the *utility* of the production with the incorrect response. Changes in utility through learning are signified by up and down arrows in this diagram.

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

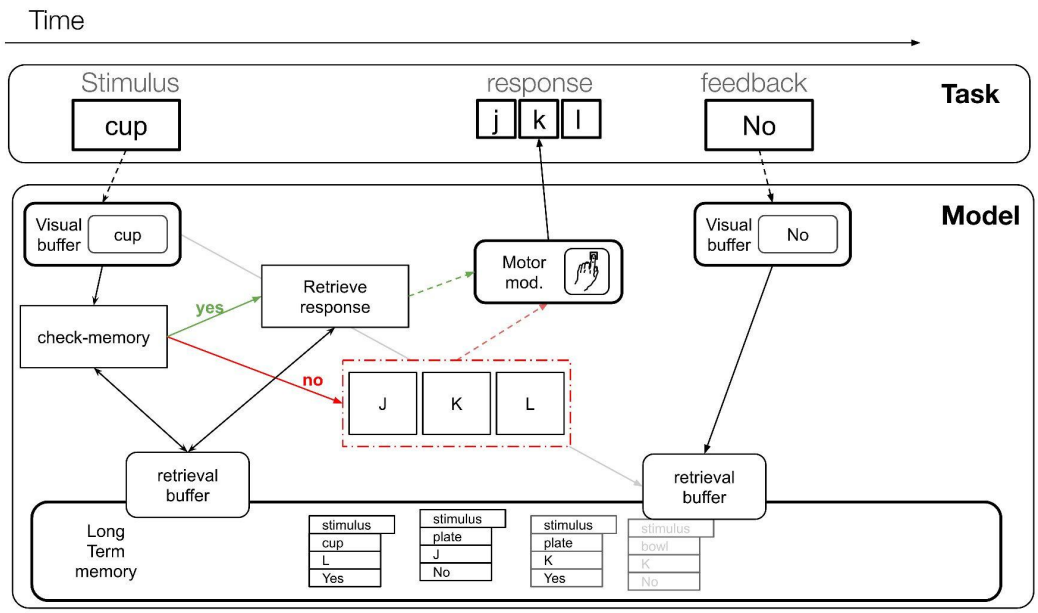

Figure 3: Overview of the declarative model as implemented in ACT-R.

(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

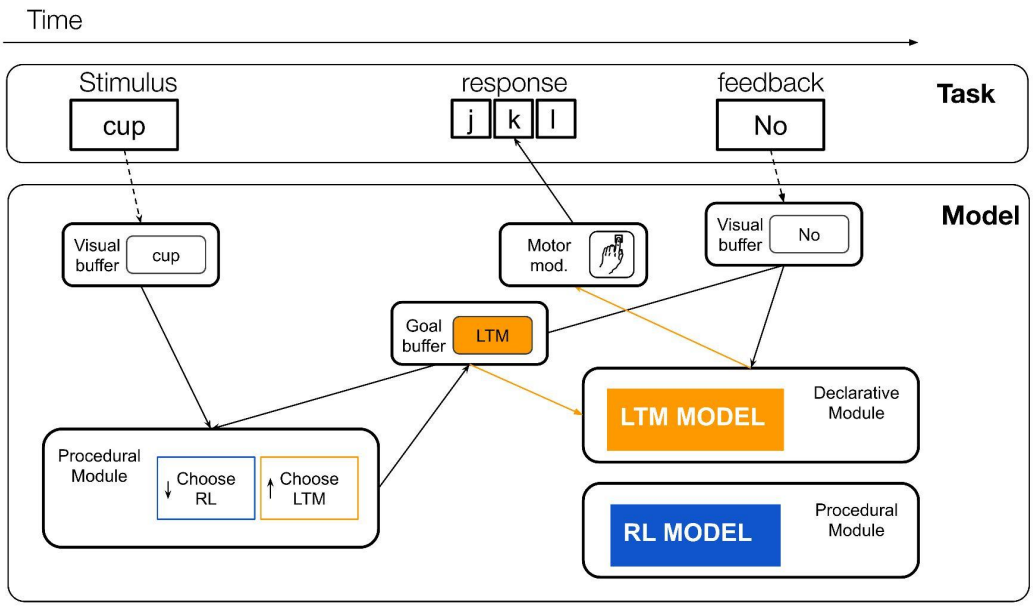

Figure 4:An overview of the integrated Meta-RL model. The meta-learning model implements a utility learning component that selects the most successful sub-model (RL vs LTM). The figure above exemplifies a scenario where the LTM sub-model was selected. Up and Down arrows in productions signifies continued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

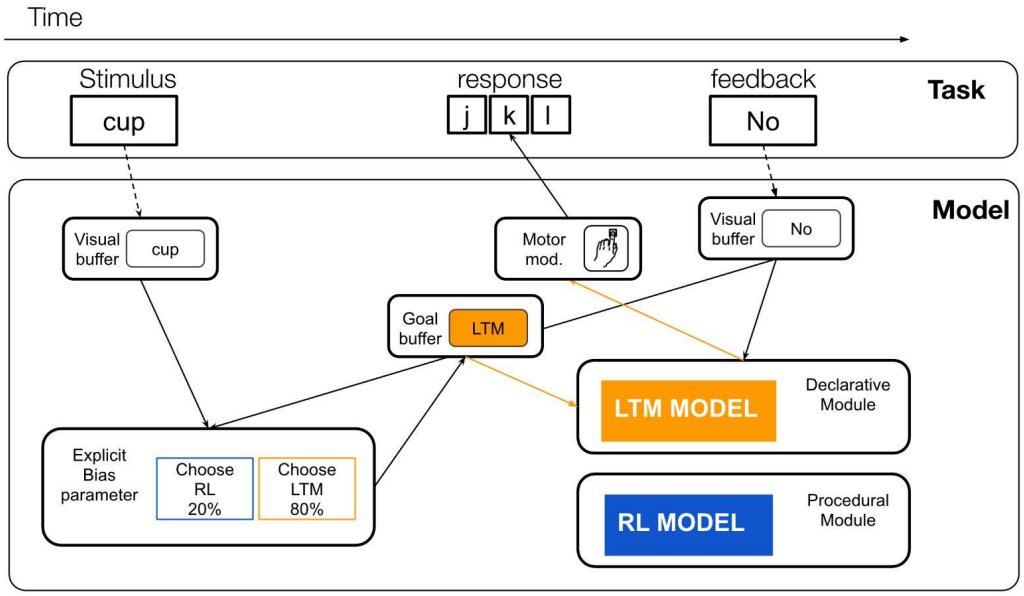

Figure 5: An overview of the integrated Biased model. The explicit Biased model does not use the procedural module utility learning but instead selects either RL or LTM at predefined proportions of either 20%, 40%, 60% and 80%. The specific example above shows a time-point where the bias parameter prefers 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
	decay rate (d)	0.3	0.4	0.5	0.6	0.7
LTM	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:

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

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:

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

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.

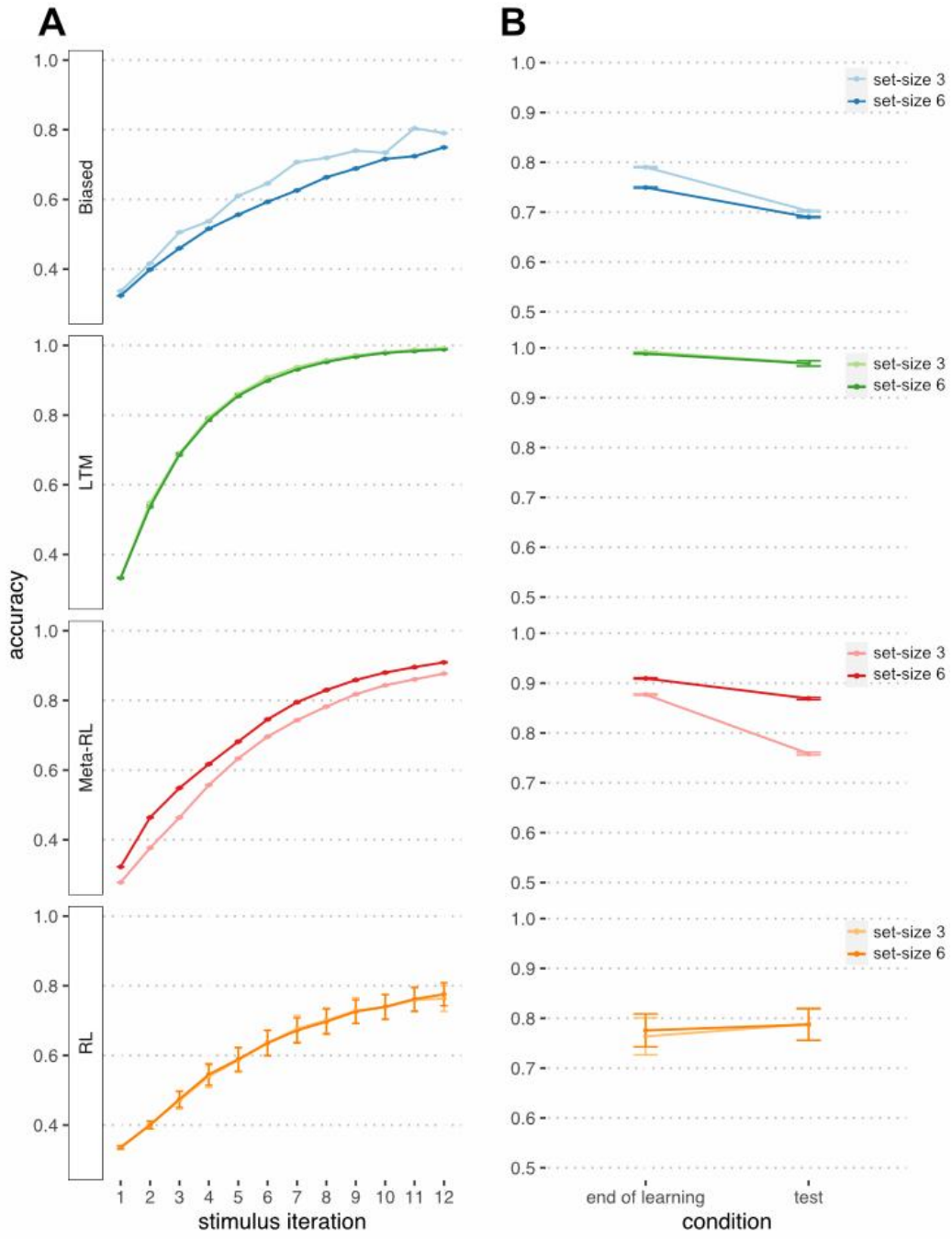

Figure 7: Model Learning accuracies for successive presentations of stimuli (A) and accuracy at end of learning, and test (B). The data points are mean accuracy averaged across all parameter-sets. Light colored lines, points and curves represent set-size 3 and dark colored points, lines and curves represent 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).

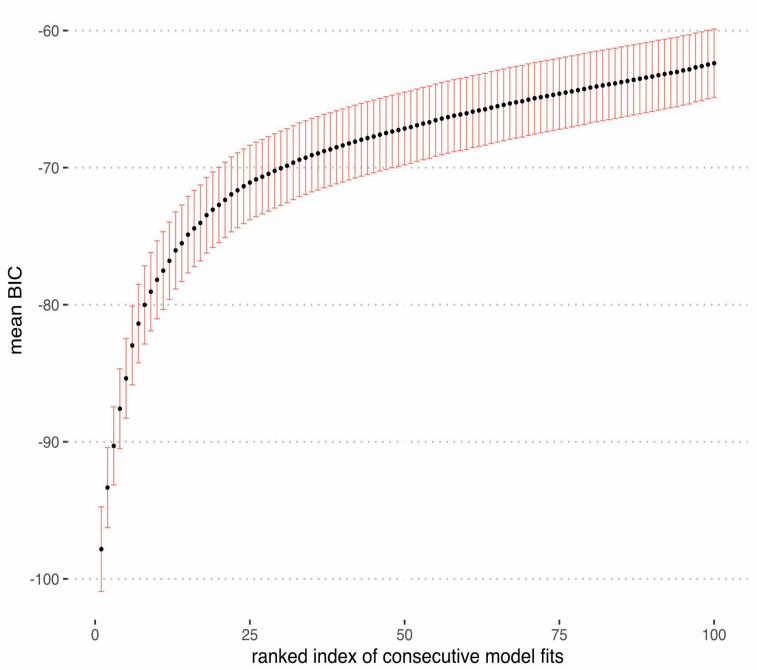

Figure 8: Mean BIC values for ranked models fits, therefore, the first data point shows the mean BIC of the best fit model, which might be different for most participants. Subsequent models could be from the same model family, with different parameter values or entirely different models. Data points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

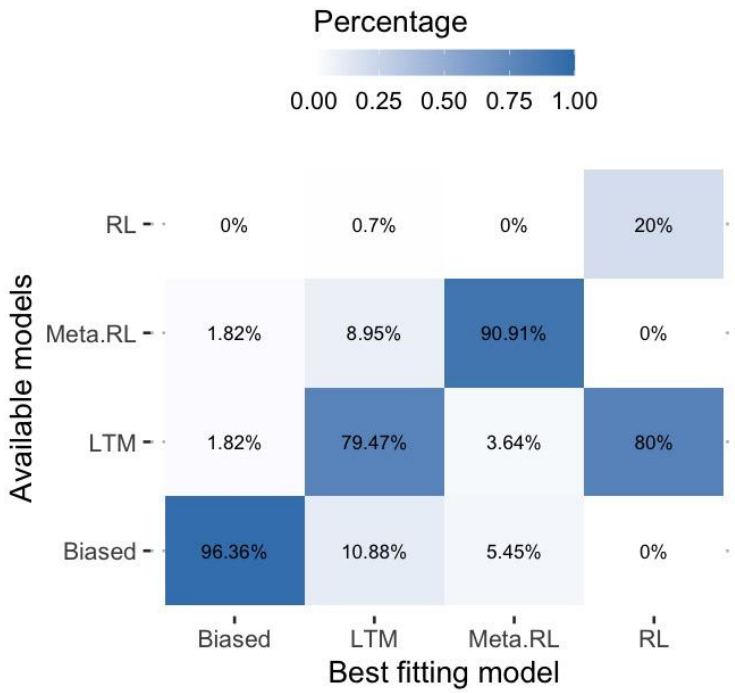

Figure 9: Confusion matrix showing proportions of fit models in the first, ranked, 10 best fit models.

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 reflects 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).

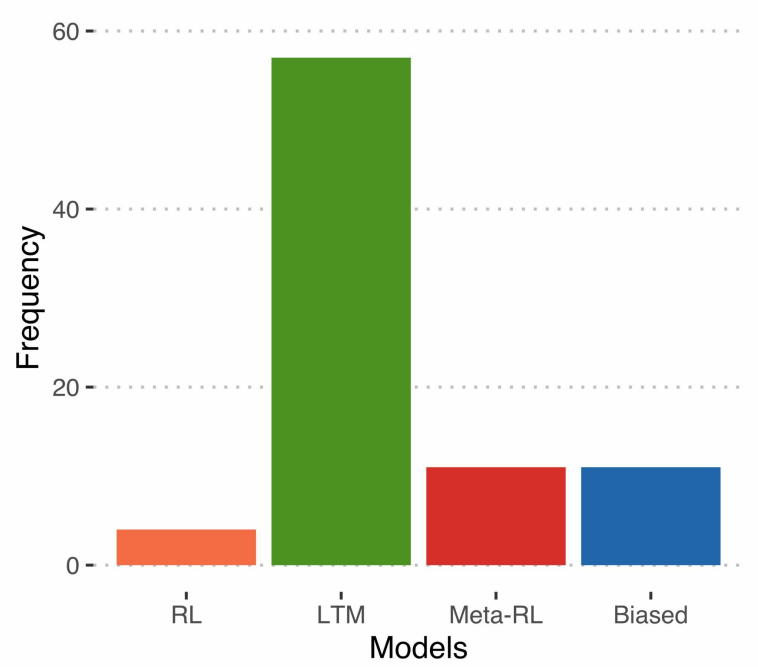

Figure 10: Counts of participants by best fit model group.

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).

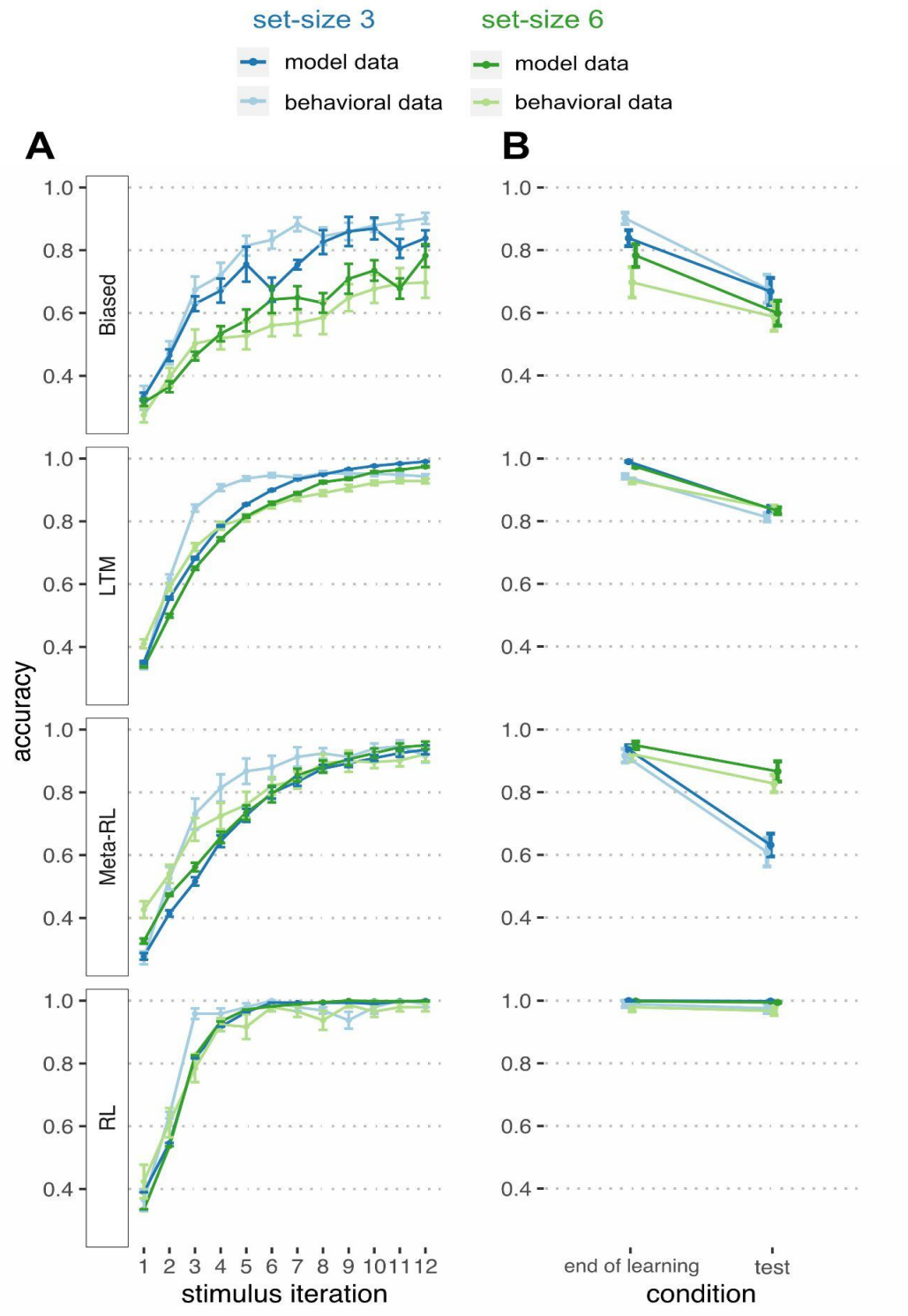

Figure 11: (**A**)Mean learning curves for set-size 3 (blue) and set-size 6 (green) objects.(**B**) Mean accuracy of learning in the last stimulus iteration, and testing. Light colors are averages of subject data that fit that model best. Dark colors are averages of model data for parameter-sets that 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

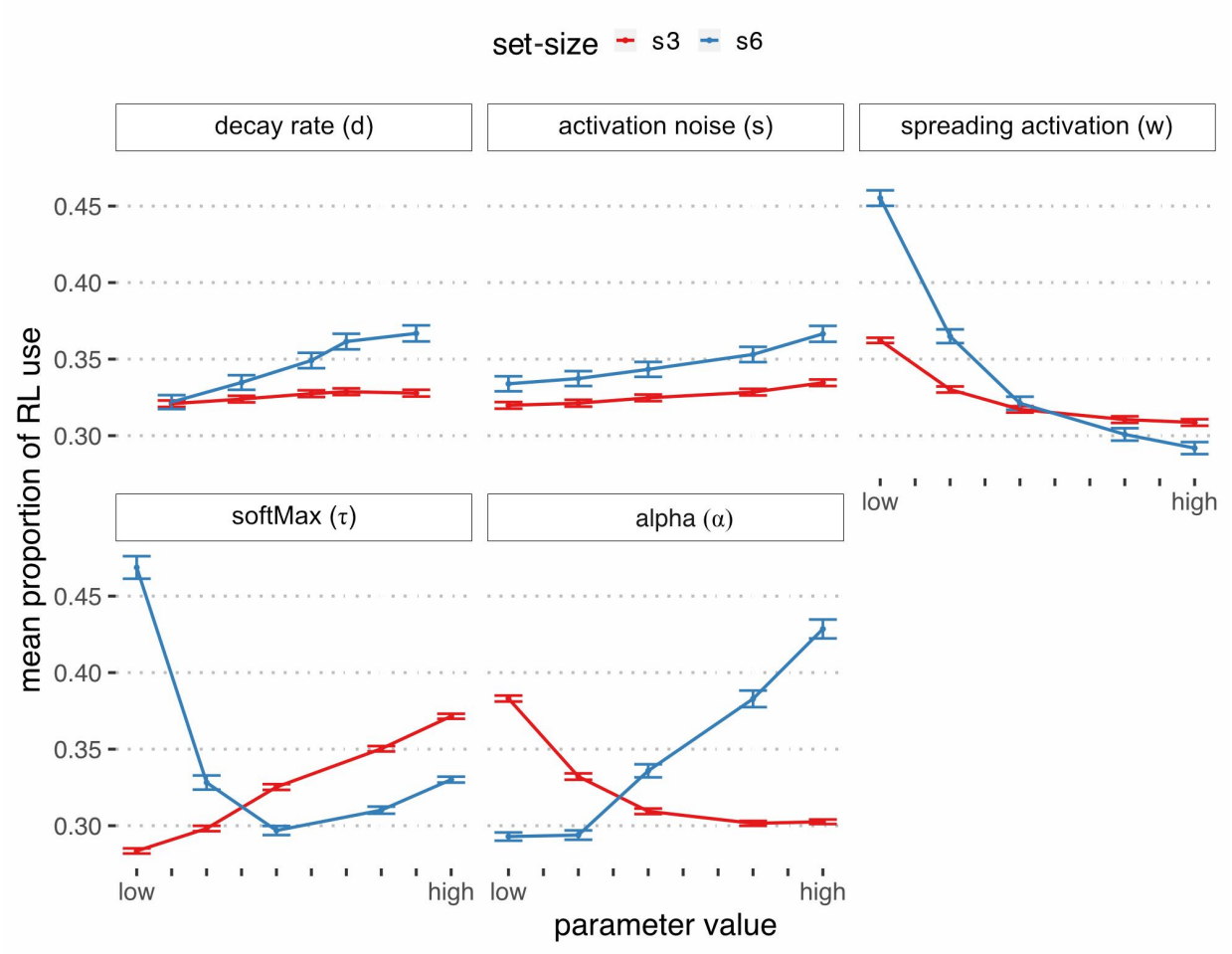

Figure 12: Model data only showing the effect of changes of parameter values on the proportion of RL used for the set-size 3 and set-size 6 blocks, in the Meta-RL integrated model. Top 3 panels are LTM parameters and the bottom 2 panels are RL parameters.

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 can not 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 imaginal-activation. In one instance, the proportion of RL use fell for both set-sizes as the imaginal-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.