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

## Abstract [150-250 words]

Complex skill learning depends on the joint contribution of multiple interacting systems: memory (WM), 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. To do so, we built four ACT-R idiographic learning models (single-mechanism RL and LTM, and two integrated RL-LTM, meta-learning RL and parameterized RL bias models) using the Collins (2018) stimulus-response association task. Different models provided best-fits for individual learners (LTM: 73.5%, RL: 2.4%, meta-learning: 10.08%, bias-RL:13.25% of participants), which suggests that irreducible differences in learning and meta-learning strategies exist between individuals. Models predicted learning accuracy and rate, and testing accuracy for subjects in their respective groups.

Keywords:

Note: I’m going to rewrite the Abstract once the paper is finished.

***Todo:***

***Changelog:***

* Reverted to using declarative memory instead of Long-Term Memory following Chantel’s feedback.

## Idiographic computantional models reveal individual differences in learning strategies

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 learning and memory mechanisms including working memory, reinforcement learning, and declarative memory (Stocco et al., 2010). Though a considerable amount of research has investigated how task characteristics drive these mechanisms during learning (e.g., Collins & Frank, 2012; Poldrack et al., 2001), less work has been devoted to understanding how and when they may be differentially deployed in different learners. To examine these individual differences in learning mechanism use, we built four models, using the Adaptive Control of Thought - Rational (ACT-R) cognitive architecture, which employed different learning strategies to complete Collins’ Reinforcement Learning Working Memory task (RLWM task; Collins, 2018) and fit them to individual learners.

Collins’ RLWM task was chosen because it provided a single experiment with simple manipulations to dissociate learning mechanisms. In this task, participants are asked to learn associations between images (e.g., objects, shapes, and colors) and letters by using the feedback provided. The task sought to quantify the relative contributions of working memory and reinforcement learning through two training conditions. The first condition consisted of short blocks containing only 3 images while the second condition consisted of long blocks of 6 images. 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 be overwhelm working memory capacity limitations, making the system unreliable, and would therefore be learned through reinforecement learning, which is not limited. To evaluate which system was used, working memory or reinforcement learning, this simple task also contains a surprise test after a 10-minute active break (figure 1). If working memory was used to guide responses, the learned associations would decay in the 10-minute active break and become inaccessible during the test. The stimulus-response associations learned through reinforcement learning would survive the break and the learner can behaviorally demonstrate learning. This largely aligns with what we know about the durability of reinforcement learning (e.g., Stocco et al., 2010). However, in the task’s simplicity lies a difficulty not yet considered; it is difficult to make predictions about how declarative memory and reinforcement learning might have guided actions during the testing phase using behavioral outcomes only as both of these systems would result in enduring representations associations.

It can be difficult to assess the independent contributions of different learning mechanisms behaviorally. Modelling is a robust approach to evaluating the independent contributions of these learning mechanisms (e.g., Stocco et al., 2021; Collins, 2018; Stocco et al., 2017; Daw, 2011). Collins demonstrated the relative contributions of reinforcement learning (RL) and working memory (WM) to learning the object-letter associations using a combined, interacting WM - RL model (RL+WMinteracting; Collins 2018). Collins 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 ineracting RL+WMi model fit participants’ data best compared to other, RL only and non-interacting RL+WM models (Collins, 2018).

Diagram, text

Description automatically generatedOne critical limitation of the Collins (2018) original modeling effort is that it implicitly assumes that all long-term associations between stimuli and responses are stored in a procedural, RL-based system, and, conversely, that all of 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 5-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.

To further complicate the story, Collins’ model relies on a simplified working memory system, which, in essence, is a fixed-capacity storage with fading contents. This is exactly how short-term memory was originally conceptualized by Atkinsons and Shiffrin (1968) and, while useful as a modeling tool, it is also known to be inadequate. Critically, many contemporary theories think of working memory as a process arising from the interaction between attention and the strategic retrieval of long-term memory information (Kane et al., 2001; Miller, Lundqvist, & Bastos, 2018). Collin’s modeling efforts confound the temporal axis of learning (long vs. short term representations) with the learning representation (implicit and procedural, driven by RL, and explicit, driven by WM).

Lastly, Collins (2018) does not take account of individual differences in learning strategies, working memory, and reinforcement learning. Collins (2018) uses a group-level model-fitting procedure that is tolerant of individual differences but does not try to fit different models to single subjects. Individual differences in both WM and RL are well documented and these differences impact learning outcomes, especially when the to-be-learned tasks tend to be more complex. For instance, in an individual differences study, learners with high working memory capacities were slower to learn categories where separation of items was based on multiple relevant features (DeCaro et al., 2008). This suggests that these high working memory learners tried to acquire the categories suboptimally by maintaining object features in working memory. Relatedly, Kalra et al. (2018) compared several implicit (RL-based) and explicit (WM and declarative memory based) tasks in a test-re-test experiment. They showed that there are stable individual differences in both explicit and impicit learning that produced correlated performance in the re-test sessions. However, learner performance on complex tasks like the Artificial Grammar task (Knowlton & Squire, 1996), that seemingly draw on both implicit and explicit mechanisms, were not predictable because learners likely changed strategies that disproportionately drew on explicit and implicit mechansism in the different testing sessions. Gluck (2002) has documented similar individual differences in learning strategies in the Weather Prediction Task, where different participants focused on different portions of the stimuli to learn the probabilistic categories. These studies provide a more complete view of learning and interactions between individual differnces and multiple learning mechansims.

Collins (2018) provides computational modeling evidence for interacting mutliple learning mechansims that aligns with other behavioral and neural evidence for dual systems for learning (E.g. Anderson, 1982; Poldrack, 2001; Antzoulatos and Miller, 2014). However, the study de-emphasizes the individual cognitive differences that might dictate how these systems interact and, presents an incomplete acccount of explicit learning that very likely relies on declarative long-term memory in additon to working memory (Poldrack, 2001; Schneider and Chein, 2003).

### Modeling individual differences in learning using ACT-R

To capture the interplay between RL, declarative memory, and WM in an integrated model 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 usability is learned through RL (Ceballos et al., 2020; Stocco et al., 2010). Declarative knowledge is represented as explicit memories. Explicit memories decay over time, but their 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 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 models to capture sensorimotor interactions with a task.

In the current study, we built four models to model typical learning trajectories and outcomes in a declarative memory only system (LTM model) with a variable WM analog, a RL only system (RL model) and two combined RL, WM, and LTM models (RL-LTM models). These models would allow us to uncover the potential contributions of declarative memory to the RLWM task.

## 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 (47 females, aged 18-35 years). Data were collected after receiving informed consent in one 2-hour session.

**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 reward (+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.

**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,

*Ut (p) = Ut -1 (p) + α [ R t - Ut -1 (p)]* (1)

Diagram

Description automatically generatedin which *α* is the learning rate and *Rt* is the reward given at time *t*. In our experiment, R*t* is binary and corresponds to the feedback (“Correct”, *Rt* = 1, and “Incorrect”, *Rt* = -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*. [INSERT SOFTMAX RULE EQUATION HERE]

The entire RL model is controlled by those two parameters, the learning rate *α* and the selection noise *τ*.

**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 (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 Diagram

Description automatically generatedduring learning and retrieval[AN EXAMPLE MIGHT BE GOOD HERE – ASK CHANTEL].

Figure 2: Overview of the procedural RL model as implemented in ACT-R.

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

**Integrated LTM-RL models.** Our third and fourth models (Figures 4 and 5) integrate the two single-system models into two new mulit-component RL – LTM models with differences in how a 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 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 Diagram

Description automatically generatedshort 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.

Figure 4:An overview of the integrated models. The meta-learning model implements a utility learning component that selects the most successful sub-model (RL vs LTM). The explicit bias model selects either RL or LTM at a pre-defined proportion.

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.

### 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. 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.

Table 1: Parameters for Procedural (RL) and Declarative modules (LTM) explored in grid search.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Parameter | Value | | | | |
|  | alpha(α) | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 |
| RL | 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 |
|  | activation noise (s) | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| LTM | spreading activation (w) | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |

**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 *θM*, that minimized the following function:

*M, θ = argmin BIC (Yp, YM | M, θ)* (3)

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

BIC *= n + n log (2π) + n log (RSS)/ n) + log (n) (k + 1)* (4)

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 particular individual.

## Results

### Behavioral Results

By and large, our experimental results replicated the experimental findings of Collins (2018). On average, participants’ performance improved throughout the learning phase of the experiment, as shown by a significant effect of the stimulus repetition on its response accuracy (*F*(11,984) = 405.67 p <0.001). As previously reported, stimuli in the set-size 3 condition was 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* = 0.054: *see Figure 6*). Finally, the two conditions (set-size 3 x learning/test phase) interacted (*F*(1,328)=8.14, *p* < 0.001), with greater forgetting in set-size 3 as compared to the set-size 6 between training and test (Figure 7). As noted in Collins (2018), 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, and a preservation of information for set-size 6 objects which aligns Chart, box and whisker chart

Description automatically generatedmore with a procedural memory utilization. Additionally, the superiority in speed and accuracy of stimulus-response learning of objects in the set-size 3 condition rules out Chart, line chart

Description automatically generatedreinforcement learning.

Figure 6: Accuracy across successive stimulus presentations during the RLWM task.

Figure 7: Accuracy during the test phase in the RLWM task.

### Overview of Modeling Results

Figure 8: Model Learning accuracies for successive presentations of stimuli. The data points are mean accuracy averaged across all parameter-sets. Light blue curve and points represent set-size 3 and dark blue points and curve represent set-size 6. Top-left ‘bias’ figure shows the learning trajectory for the explicit bias strategy integrated model (12,500 parameter-sets). The bottom-left ‘metaRL’ figure shows learning trajectories for the meta-learning RL integrated model (3125 parameter-sets). The top-right ‘LTM’ figure shows learning for pure LTM model (125 parameter-sets) and the final figure, bottom-right shows learning for the pure RL model (25 parameter-sets).

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 8), which suggests that different learning strategies alone produce variable outcomes. As Collins (2018) pointed out, the pure RL model predicted no difference between set-Chart

Description automatically generated with medium confidencesize 3 and set-size 6. Notably, the 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. It is difficult to understand why, but the 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. Remember that the specific parameter values influence the success rate of the sub-models (see Capturing Individual Differences section 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 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-learning model achieving close to 90% accuracy, the pure RL model around 78% and the biased model close to 70%. Only the biased model resulted in a difference of 10% in final accuracy between the two set-sizes.

### Model Fitting Outcomes

Chart, bar chart, box and whisker chart

Description automatically generatedAfter 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 was ranked 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). We found that the difference in BIC value for the best fitting and the second-best fitting model was in the positive evidence range (M= 4.76; SEM=0.451). This difference fell when comparing the 2nd with the 3rd best fitting models (M=3.28; SEM= 0.357) and 3rd with 4th (M=2.75, SEM=0.364) (figure 9). 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 (*P(M|D),* Kass and Raferty,1995). When split by model type, cases where the LTM model fit participants best demonstrated the strongest evidence against the 2nd best fit model compared to the other models. This suggest that the LTM model overall has the highest likelihood of capturing learning behavior in the RLWM task, even while it is not the model with the least number of parameters.

Figure 9: Differences in BIC for ranked consecutive fit models. Data points show mean BIC difference with standard errors. Red lines show the 6 to10 (strong evidence), 2 to 6 (positive evidence) and 0 to 2 weak evidence ranges.

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, best-fit models. We found that the best-fit model was selected on average 8.34 times out of 10. The best fit model was selected 91.8% of the time for the biased family of models, 84.4% for LTM, 92.2% for meta-RL and 65% of the time for the two RL fitting participants. 4 participants (1 from LTM and 2 from meta-RL and 1 from RL) had second best-fitting models that came from a different model family.

### 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 quantifying variability in learning: first, comparing groups of subjects categorized by best-fit model family and, secondly, predicting learning behavior from individual parameter estimations. 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 = 61), followed by the biased model (n = 11) and the meta-RL model in third place (n = 9). Only two participants best fit the pure RL model. The group differences are shown in figure 11.

Chart, bar chart

Description automatically generated

Figure 11: counts of participants by best fit model.

Secondly, we were interested in capturing the participants’ specific behvioral features during learning and test such as their learning rate, forgetting after break, and difficulty associated with the increase in the number of objects to learn.Graphical user interface

Description automatically generated The models’ performance faithfully matched the group-level behavioral performance in the stimulus-response accuracy in both the set-size and learning-testing phase conditions. We found that the objects in the set-size 3 blocks were learned faster and with higher accuracy than set-size 6 objects (see behavioral results section above). The subset of models that fit participants best from the LTM and explicit-biased models predicted the observed higher set-size 3 learning rate (LTM: *W* = 3139, p < 0.001; biased: *W* = 99, *p* = 0.0104) and accuracy at end of learning (LTM: *W* = 2917, *p* < 0.001; biased: *W* = 113, *p* = *p* < 0.001) compared to set-size 6 (figure 11*,biased* and *LTM* panels) as determined by a Wilcoxon rank-sum test. The subsets of models that fit participant data from the meta-RL and RL only models predicted similar learning rates between the two set-sizes, which was not reflected in the observed data (meta-RL: *W*=31, *p*=0.426; RL model has only 2 data points so statistical testing was not performed).

Figure 11: Mean learning curves for set-size 3 (blue) and set-size 6(green) objects. 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 subjects in that model group.

We have shown that, on average, set-size 6 objects were remembered better than set-size 3 objects after the break, but the meta-RL model has more faitfully captured the difference between the two block sizes, compared to the other models (figure 12B and 12D). It should be noted that best-fitting Collins model (RLWMi) predicted the higher forgetting in set-size 3 compared to set-size 6 during test after the break, but we found varying results. The behavioral data from LTM group (our largest group of learners) and biased integrated model shows the same amount of forgetting during test for both set-sizes which defies the WM – RL dichotomous view forwarded by Collins, 2018, while the meta-learning group follows that pattern of more preserved set-size 6 memory during test (Figure 12D). Even then, the meta-RL group had 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 test.

We observed different rates of difficulty, the difference between the learning trajectories for set-size 6 and set-size 3, among the behavioral data that were grouped by the best-fit models

|  |  |
| --- | --- |
| Chart, box and whisker chart  Description automatically generatedA | Chart, box and whisker chart  Description automatically generatedB |
| Chart  Description automatically generatedC | A picture containing text  Description automatically generatedChart, line chart  Description automatically generated D |
| Figure 12: Characterization of model fits for learning features of interest. A) differences of the set-size 3 and set-size 6 learning curves across the 12 stimulus iterations. B) Difference in accuracy from end of learning to test. C) Learning rate (slope) for the first 6 stimulus iterations. D) End of learning and test accuracy. | |

(RL model was excluded since it contained only 2 data points) as shown by a Kruskal-Wallis 1-way ANOVA test (χ2(2) = 23.74, p <0.001) (Figure 12A); differences between meta-RL and LTM were not significant p=0.830, in a Wilcoxon pairwise test, Bonferroni corrected for multiple comparisons. The models on the other hand predicted significant differences between the groups (χ2(2) = 36.58, p <0.001) and pairwise comparisons also showed that all combinations of comparisons except for RL were significant at an alpha of 0.001. Lastly, figure 12C further quantifies the learning rate differences among the different groups which were well captured by the models.

### Parameter analysis

Assessing the parameters’ effects on learning outcomes is difficult as the relationships among parameters is complex and determined by the specific formulae in ACT-R. However, some parameters might have linear and interacting effects on learning outcomes. To reiterate, the model learning was influenced by a maximum of 6 parameters ranging on 5 values (Table 1). The two integrated models, meta-RL and explicit biased model utilized the 5 parameters described in Table 1, along with an additional bias parameter (*β*). In the biased model, the bias parameter defined the proportion of RL vs LTM to use throughout the learning and test phases but *β* was measured for the meta-RL model by determining the proportion of RL used by the meta-learner throughout the learning phase.

### A wide 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 35 parameter-value sets across all four models, out of 15,865 possible sets, described all 83 subjects. Subjects who fit the most popular model, LTM, that fit 61 subjects, surprisingly, were described only by 14 parameter-value sets for the 3 LTM parameters out of the possible 125 combinations (spreading activation, retrieval noise and memory decay rate). The biased model was the most diverse at 11 parameter sets for the 11 subjects in that group (out of a possible 3125 combinations). The meta-RL model had 8 parameter-value sets for the 9 subjects and, lastly, there were 2 RL best fitting combination of parameter values for the alpha and rule selection noise parameters. This demonstrates that even within the four groups of participants fit by the different models there are notable individual differences to be captured. ~~The variability of parameter-sets within the integrated models, suggests that individual differences persist even at the level of meta-learning, or deciding which learning mechanisms to apply.~~

Across all subjects that fit LTM strategy containing models, on the memory decay parameter had a negative skew:75% fell on the highest level in the range explored and the next 20% were characterized by the second highest value of decay rate. For the retrieval noise parameter on the other hand, fit paramters were, relatively more uniformly sampled from the range of values explored with a less pronounced negative skew (only 35% of fit subjects had the highest decay rate). 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 50% of subjects fell onto the lowest value 0.1, followed closely by 27.1% for the second lowest parameter value.

Regarding the two RL parameters, RL noise and learning rate, both parameters are slightly skewed but in opposing directions. It should be noted that the interpretation of these correlations is limited due to the small size of the samples in all but the LTM only model. The noise parameter is positively skewed (40% of fit subjects falling on the lowest values of 0.1).

There is not sufficient variability in these data to estimate the isolated effects of the parameters using linear methods. But it can also be taken as evidence that perhaps, single parameters in isolation may not have large driving effects on learning outcomes; 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 bias in the meta-RL model is explored in detail in the next section.

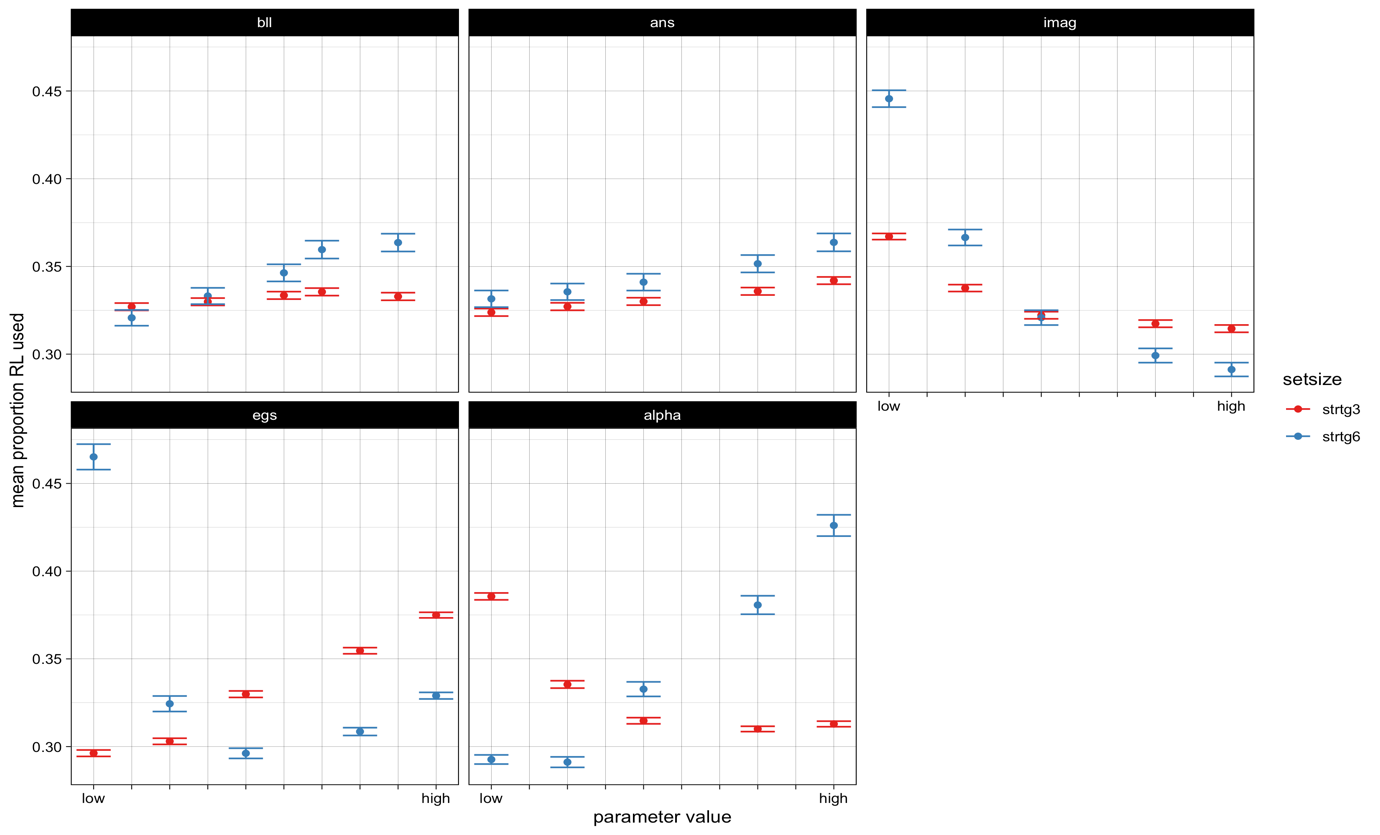
The 6th parameter, learning bias, was pre-defined in the biased model (figure 5) and had the values 20%, 40%, 60% and 80% probability of RL use. We found that 9 out of 11 subjects that fit this model used RL only 20% of the time, adding to the growing evidence of general preference of a declarative strategy that we have uncovered, shown by the large group of subjects that fit the pure LTM model. 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 parameter, here, was measured at the end of learning by taking the mean number of times the RL sub-model was deployed. This RL meta-learner selected different proportions of RL and LTM for the two set-sizes. 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 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 9 of our 83 participants fit this model best but we observed that the estimated bias towards RL in the set-size 6 increased significantly, to around 50% of the time, compared to set-size 3 (set-size 3: M = 0.3; SEM= 0.05; set-size 6:M = 0.47, SEM = 0.15). It would seem that RL was prefered, in general to learn the set-size 6 blocks, which aligns with Collins’ (2018) observation, but a very small subset of our participants used this distinct different preference in strategies to deal with the two different set-sizes.

Figure 13: 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 on the proportion of RL sub-model used in the meta-learning RL integrated model. Parameters: **alpha**:RL learning rate, **egs**: RL selection noise, **bll**: memory decay rate, **imag:** attentional spreading activation weight, **ans:** activation noise.

## Conclusion

## We implemented four, idiographic ACT-R models and demonstrated that learners use different learning strategies (RL vs declarative memory) to engage with even simple stimulus-response learning tasks, with some adapting their learning strategy during the task (more likely to fit the meta-learning model), and some maintaining a bias towards one learning system (more likely to fit the explicit-biased model). We further demonstrated that task learning heaviliy skews to a declarative memory strategy. Therefore, our results highlight the value of the idiographic approach, the importance of declarative memory in learning, and new insights as to why some may prefer some learning strategies.

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 — an approach that has recently gained prominence in cognitive neuroscience (Ceballos, Stocco, & Prat, 2020; Daw, 2011). The original study, Collins (2018), found a best-fit model for the entire cohort of subjects which does not explain how or why different individuals might use different strategies. The success and prominence of RL theory in neuroscience has led to probably overlooking how much individuals rely on declarative strategies in learning simple response associations 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.

We found that a large majority of our subjects prefered a declarative memory strategy to learn both the set-size 3 and set-size 6 blocks (73% of subjects fit the LTM model best) and there was minimal forgetting after the break for both set-sizes. This pattern of results is a departure from those obtained by Collins (2018), where part of their evidence depended on the different levels of forgetting between the set-size 3 and set-size 6 to explain that the smaller object size blocks were learned with working memory and the larger set-size group must have been learned with RL since it survived the break relatively intact. In addition to the results from the LTM only model, the explicit bias model showed that the declarative memory strategy was highly preferred, at 80%, over the RL portion. We demonstrated that learning could also occur through declarative memory with still robust effects and are also 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).

Part of our results resembled those obtained by Collins (2018), and demonstrated that some learners prefer to use a mixture of strategies to learn the different sized blocks. Additionally, and importantly, our parameter value estimates aided us in understanding, in part, why some of these subjects might choose different strategies, and arbitraring between RL and declarative memory both of which are robust to distracting breaks and difficult to isolate behaviorally. Subjects who fit the meta-RL mixed model forgot more images from the set-size 3 than set-size 6 as shown in Collins (2018). This pattern of forgetting is consistent with assumptions that most of the set-size 3 objects were learned with WM, thus facilitating performance by direct reading of the associated response from a short-term buffer. The set-size 6 objects, on the other hand, were learned with either RL or Declarative memory. Here, again, our idiographic approach lends a hand to further understanding. The meta-RL model, which has an RL meta-learner was designed to capture a likely situation where learners would depend on their relatively recent success with learning and adjust strategies. We found that these subjects used a RL strategy only about 30% of the time to learn the set-size 3 objects and about 50% of the time to learn the set-size 6 objects which aligns with our learning flexibility and adjustment hypothesis. Examinging the model data, we found that the meta-RL learner de-emphasized either the declarative LTM or RL sub-model depending on their relatively poor performance, which was heavily influenced by the combination of parameter values. For instance, the proportion of RL sub-model use fell for both set-sizes as the imaginal-activation injected into the LTM model increased; RL use was highly anticorrelated for the two set-sizes, with a large RL preference for the set-size 6 block as the learning rate parameter increased. We can venture to assume that individual learners similarly have varying intrinsic cogntive properties that affects learning strategy choice. To the best of our knowledge, this is the first study to report such findings.

A number of limitations must be acknowledged. First, the number of models we explored is still limited. Second, and most importantly, the size of the parameter space that was explored was extremely small and coarse, limiting our ability to capture nuanced individual difference. Both of these limitations will need to be overcome in future research and are currently limited by computing power. We are leveraging the use of cloud computing, as suggested by one of our reviewers, to search a wider range of parameter values. Third, the BIC model selection method based on residual sum of squares (equation 2) were designed for linear-models and do not necessarily hold for arbitrarily complex models like ACT-R (Stocco, in preparation) ­— we are implementing a log-likelihood approximation to make more sound model fits.

These limitations notwithstanding, a number of important points need to be made. The first is that individual differences do matter and, as it is becoming increasingly apparent, group 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.

More importantly, it was found that the principle that different individuals fit different models also applies to higher-level models. In our case, the two “integrated” models were found to better fit different participants, To the best of our knowledge, this is the first study to report such findings.

A third and related point that needs to be made is that, while models do matter, the specific type of modeling approach that is used matters even more. It would have not escaped the attentive reader that, while our empirical results largely mirror those of Collins (2018), our conclusions do not. This is mostly due to the fact that our choice of modeling paradigms was different, and carries different assumptions about the cognitive system. Consider the difference in learning between set-size 3 and set-size 6 conditions. Collins’ (2018) explanation is that set-size 3 items are more likely to be still in working memory during learningOur explanation is that participants probably relied on different learning systems, LTM vs. RL, for the two sets of stimuli. Because the space of possible models is so large, it is practically impossible to empirically decide on this matter. For this reason, we advocate for developing idiographic (i.e., individual-level) models within an integrated cognitive architecture, so that the different models are more clearly comparable and benefit from a common, well established set of constraints (which seems to be evolving towards a consensus: Laird, Lebiere, & Rosenbloom, 2017). By doing so, we believe we have put this research on a better footing for future developments.

## What does it mean?

These results suggest that individual differences are prevalent, even during simple stimulus response learning tasks, and that idiographic modeling techniques can be used to capture these differences (reminder of results that back this up...) then  
  
"Using the ACT-R framework, our results suggested that long term declarative learning mechanisms better accounted for the learning behaviors than RL" (or that it's an alternative characterization worth considering). Now talk about how this is different than how collins described it and provide your evidence... How many people fit LTM, How well did it account for the data, etc.

Chantel: I think the first paragraph of the conclusion should read like an abstract... Idiographic models should be in the first sentence, and then you should NOT repeat what we did but instead say something like... each of four models that differentially relied on declarative versus procedural learning fit some subset of our participants best... Kind of emphasize why what we did highlights the utility of the idographic approach

**References**

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.

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).

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W.(2001). A controlled-attention view of workin-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.

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.

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

Stocco, A., (in preparation). Log-likelihood estimations for ACT-R models.

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.