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

Theodros Haile (theodros@uw.edu)
Department of Psychology, University of Washington
Campus Box 3515525, Seattle, WA 98195 USA

Chantel S. Prat (csprat@uw.edu)
Department of Psychology, University of Washington
Campus Box 3515525, Seattle, WA 98195 USA

Andrea Stocco (stocco@uw.edu)
Department of Psychology, University of Washington
Campus Box 3515525, Seattle, WA 98195 USA

Abstract

In humans, learning is a complex phenomenon that depends on the joint contribution of multiple interacting systems, most notably memory (WM), long-term memory (LTM) and reinforcement learning (RL). There are vast individual differences in learning mechanism deployment. It is also, often, difficult to assess, through behavioral measures, the relative contributions of these systems during learning as well the specific strategies individuals rely on in performing a task. Collins (2018) put forward a working memory-reinforcement learning combined model that addresses these issues within a simple domain, but largely ignores the long-term memory component. In this project, we built four (two single-mechanism RL and LTM, and two integrated RL-LTM) idiographic learning models based on the ACT-R cognitive architecture. We aimed to examine individual differences and estimate parameters that could explain preferential use of learning mechanisms using the Collins (2018) stimulus-response association task. We found that different models provided best-fits for individual learners with more variability in learning and memory parameters observed even within the best fitting models. Our conclusion is that irreducible differences in learning and meta-learning strategies exist within individuals even within relatively simple tasks, and that model-based approaches are necessary to characterize and explain behavioral data.

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

Introduction

Individual differences in the ability to learn new associations are foundational to most measures of aptitude— a construct that describes the readiness with which one can acquire a complex skill. But even basic associative learning paradigms, like stimulus-response mappings, have been shown to rely on a mixture of learning mechanisms including working memory, reinforcement learning, and long term memory (Stocco et al., 2010). Though a considerable amount of research has investigated how task characteristics drive these mechanisms during learning (Collins & Frank, 2012), less work has been devoted to understanding how and when they may be deployed differently in different learners. To examine this, we built two single-mechanism and two multi-system stimulus-response learning models using the Adaptive Control of Thought - Rational (ACT-R) cognitive architecture, and used them to examine individual learning mechanisms for the same learning task. Specifically, Anne Collins' Reinforcement Learning Working Memory task (RLWM task: Collins, 2018) was used as the task paradigm .

It can be difficult to assess the independent contributions of these learning mechanisms behaviorally. Modelling is a robust approach to evaluating the independent contributions of these mechanisms (Collins, 2018). This method further allows us to estimate individual parameters that would give us insight into the cognitive properties that resulted in different forms of skill acquisition (Daw, 2011). We adopted the RLWM task because it provided a single experiment with simple manipulations to dissociate learning mechanisms.

But in the task's simplicity lies a difficulty: long-term memory and reinforcement learning guide actions and responses that are nearly indistinguishable in the context of the task using behavioral outcomes only. In the RLWM task, participants are asked to learn associations between images (e.g. objects, shapes, and colors) and key responses through trial-and-error with feedback. The task, as designed by Collins, sought to quantify the relative contributions of working memory and reinforcement learning through two training conditions over 14 blocks–a working memory, resource-sparing, 3-image condition for 8 blocks and a resource-intensive, 6-image condition for 6 blocks. After training, participants performed an unrelated, 10-minute distractor task followed by a surprise test block. Collins et al. expect that that the 3-image associations, learned quickly through working memory, should not be remembered after the distracting break, whereas the 6-image associations, acquired through reinforcement learning, should be retained after the break and demonstrated during the test phase. This largely aligns with what we know about the durability of reinforcement learning (Stocco et al., 2010). Collins has demonstrated that

learning object-letter associations most probably occurs through the interaction of Reinforcement learning (RL) and Working Memory (WM) using a combined, interacting (RL+WMi) model (Collins 2018; Collins & Frank, 2012). They hypothesized that the fast-learning (high learning-rate) WM resource, which is limited in capacity and decays rapidly, represented by a decay parameter, cooperatively interacts with the RL portion of the model, directly influencing the computation of the reward prediction error. This model contributes less to reward prediction error when the set size is high. This model fit participant data best compared to other, RL and non-interacting RL+WM models (Collins, 2018).

One critical limitation of Collin's 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). Instead, our replication of the experiment shows that participants have also used their long-term declarative memory. Upon completion of the main task, participants in our study were also asked to answer the open-ended question, "*Do you recall using a specific strategy to learn the images?"* A substantial number of them reported, for instance, relying on colors, names, or other salient features of the stimuli to remember the corresponding responses. Many answers followed the common pattern "*Pictures 'A' and 'B' shared an attribute and were both associated with the keyboard response 'V', so they were grouped together".* An informal evaluation of these responses lent a trickle of confidence to the use of a possible LTM strategy, as well as the fact that participants seem to explicitly control they learning strategies Additionally, we have observed clear individual differences in learning as well as demonstration of learned associations in our subjects that stray away from the WM-RL dichotomous view of learning. For instance, a proportion of our subjects learned quickly in both object-set conditions, suggesting working memory use, but also showed that learned associations prevailed after the 10-minute break (Figures 4 and 5).

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, 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). In essence, 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).

The ACT-R Cognitive Architecture

To capture the interplay between reinforcement learning, long-term memory, and working memory within an integrated model, we decided to follow an alternative approach and build 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 cognitive mechanisms. In ACT-R, knowledge is represented in two possible formats, procedural and declarative. Procedural knowledge is represented as procedural rules, is identified with the basal ganglia, and is learned through reinforcement learning (Stocco, Lebiere, Anderson, 2010; Ceballos, Stocco, Prat, 2020). Declarative knowledge is represented in 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, and includes multiple models to capture sensorimotor interactions with a task.

In this study, we built four models to model typical learning trajectories and outcomes in a declarative learning, LTM only system with a variable WM analog, a reinforcement learning system and combined RL, WM and LTM models. These models would allow us to test if the RLWM task can potentially be performed using declarative memory. Further, by exploring a range of parameters for learning rate (α), RL noise (τ), working memory (Imaginal-activation), memory retrieval noise and decay rate, we could estimate individual parameters and establish a link to the differences that amount to varied deployment of learning mechanisms.

Materials and Methods

Participants. 83 undergraduate students from the University of Washington participated in this experiment. All participants were monolingual English speakers recruited through the UW Psychology subject pool (47 females, aged 18-35 years). 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 key-press of either 'C', 'V' or 'B' to the displayed images. In half the blocks, participants have to learn to associate key-presses with three unique images, presented 12 times in random order and in the other half with 6 unique images each presented 12 times within the block. 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 3- and 6-set learning blocks. Stimulus presentations and data collection were done in MATLAB (mathworks.com).

Computational Models

All of the models experienced the same experimental set-up — 2 learning blocks of 3 and 6 objects respectively, a 10-minute break and a test phase without feedback.

Reinforcement Learning Model. The first model (Figure 1) most closely adheres to Collin's 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 bases 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 procedural/RL model is controlled by two parameters, the learning rate α and the selection noise τ.

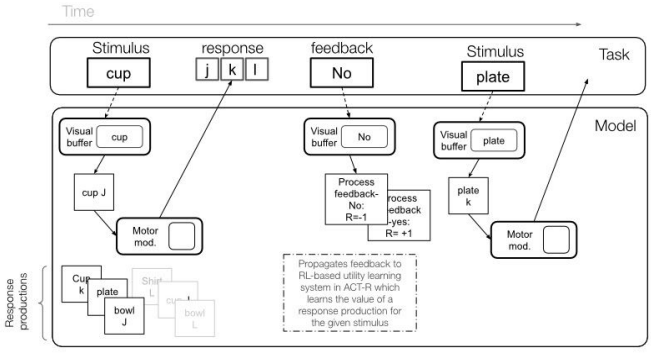

Figure 1: Overview of the procedural RL model, as implemented in ACT-R.

Declarative Learning Model. In lieu of Collins' pure WM model, we developed a *declarative* model (Figure 2), 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 response to the current stimulus that had resulted in a correct response. If such a memory is found, the same response is used. If no memory can be found, the model makes a random response. The outcome of this response to the current stimulus are then memorized. Although this model is computationally simple, ACT-R allows for a sophisticated control of the memory management processes through three parameters: (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 during learning and retrieval.

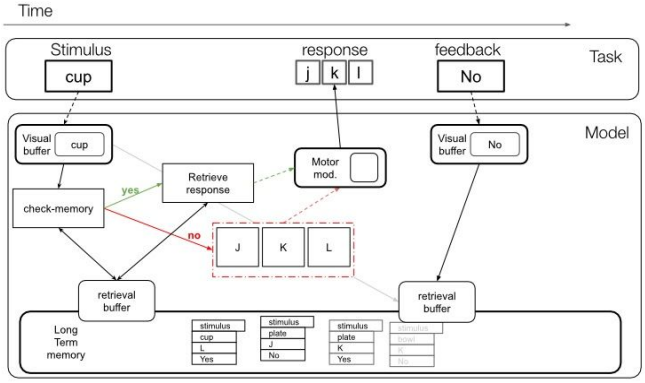

Figure 2: Overview of the declarative model, as implemented in ACT-R.

Integrated LTM-RL models. Our third and fourth models integrate the two single-system models into one model. 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, perhaps more elegant, solution was to have a reinforcement learner that learned the best strategy given the specific set of parameters. 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 second integrated model 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. This 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 has a preference for declarative learning, it would persist in trying to memorize stimulus-response associations even when switching to a RL strategy would be more convenient.

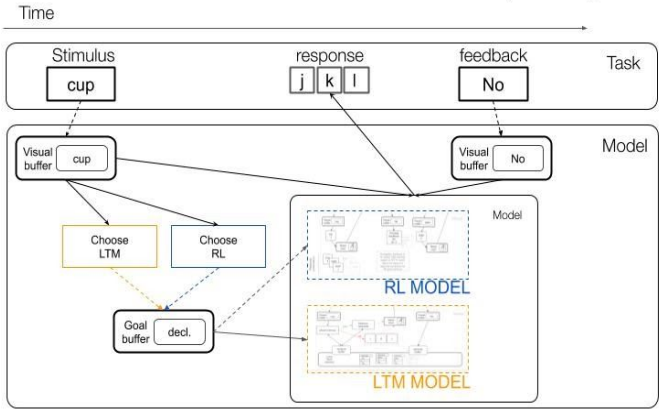

Figure 3: Overview of the two "integrated" models that employ both RL and declarative learning. The two models differ only in how they arbitrate between the two strategies.

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. A full description of parameters and the range of values that were manipulated is given in Table 1.

Table 1: Model parameters manipulated in the simulations

Parameter	Meaning	Values
α	Learning rate in RL	0.10, 0.15, 0.20
τ	Procedural rule selection noise	0.2, 0.3, 0.4
d	LTM decay rate	0.4, 0.5, 0.6
s	LTM activation noise	0.2, 0.3, 0.4
W	Spreading activation (Working memory capacity)	1, 2, 3

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 matched to a particular model M and set of parameter values θ_M, that minimized that following function:

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

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:

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

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 means accuracies associated with the presentations of each individual stimulus (12 for Set3 and 12 for Set6), 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 equally well the same data, 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). This is shown in Figures 4 and 5, which illustrate the average performance of participants across the learning phase (Figure 4) and a comparison of the end of the learning phase vs. the test phase (Figure 5) of the task.

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 Set3 condition was generally learned sooner and better than those of Set6. Finally, the two conditions interacted across learning and test phases [$F(1,328)$, $p < 0.01$], with learning for Set3 being more likely to decline from the end of the learning phase to the test phase.

As noted in Collins (2018), these group-level results strongly suggest that individuals use a mixture of declarative and procedural strategies. This is shown by the effects of the test phase (which suggest a decaying of information over time, possibly compatible with declarative memory) and by the superiority of the Set3 condition during learning (which rules out RL).

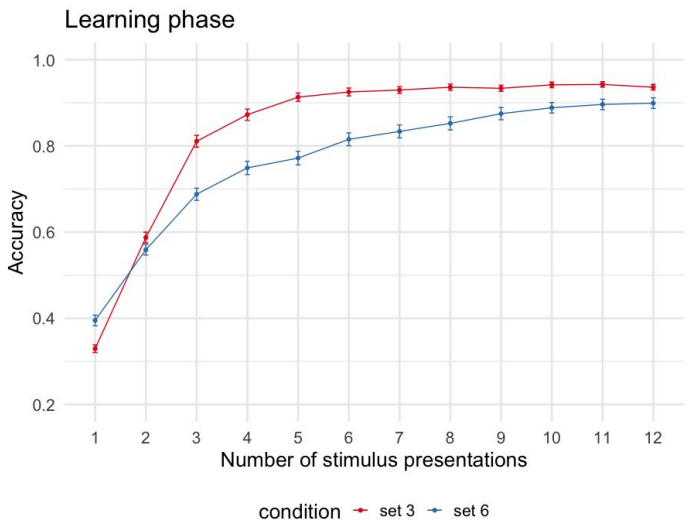

Figure 4: Accuracy across successive stimulus presentations during the RLWM task (Collins, 2018).

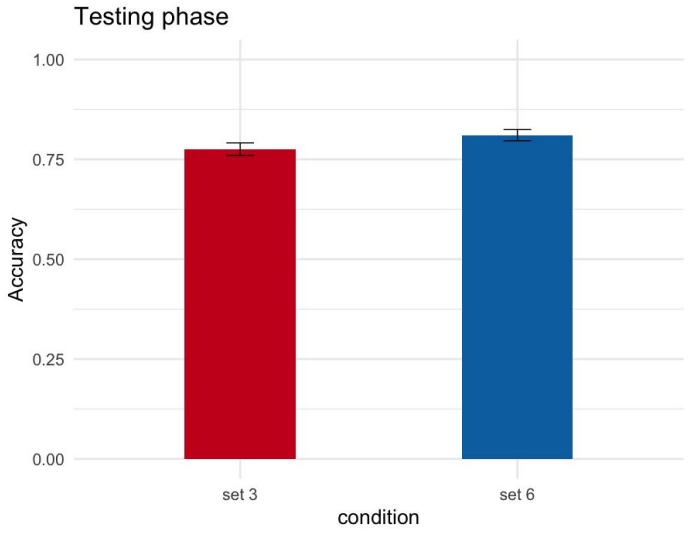

Figure 5: Accuracy during the test phase in the RLWM task (Collins, 2018).

Overview of Modeling Results

To give an idea of the general behavior of the four models, Figure 6 illustrates the mean performance of each of the four models during the learning phase. Although this data is averaged over all parameters and thus obscures the considerable variability across models (much like the group data in Figure 4 obscures the variability within subjects), it clarifies two important points. First, all of the four models, in general, capture the group-level learning rate. Second, even within the variability entailed by the different parameters, the models do predict different trends. As Collins (2018) pointed out, the pure RL model predicts no difference between Set3 and Set6. Notably, the pure LTM model also predicts no difference between the two sets, at least within our set of LTM parameters. The mixture models, however, do predict differences between the two conditions, with the difference being stronger for the explicit, biased meta-learning model. This is a side effect of the model using different strategies for Set3 and Set6 stimuli.

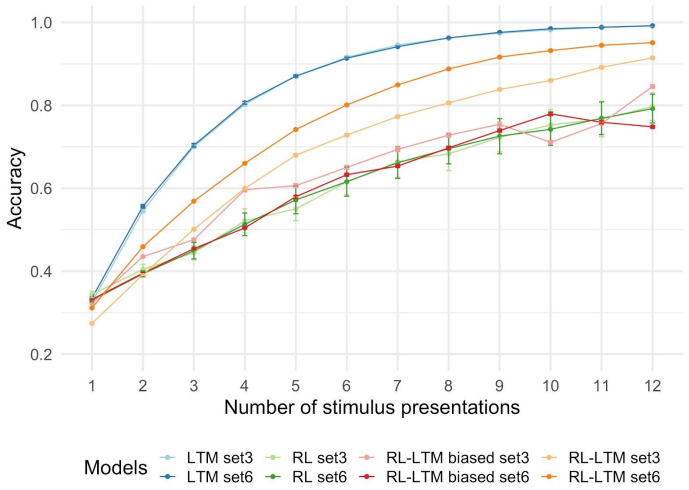

Figure 6: Learning trajectories for Set3 and Set6 stimuli for the four models.

Model Fitting Procedure

After examining the behavioral results, each participant was matched to an ideal model using the BIC criterion minimization procedure described above.

The results of this model fitting procedure yielded somewhat different results than the original study. We did not find that one model outperformed the others reliably. Rather, we found that different models steadily fit different subsets of participants (Figure 7). This was true even when, as in the case of integrated models, they effectively included the basic models as particular cases. In principle, this could be due to the fact that the BIC procedure does penalize more complex models.

Importantly, individual subgroups emerge even within the *integrated* models, suggesting that individual differences persist even at the level of meta-learning, or deciding which learning mechanisms to apply.

Conclusion

This study has used computational models to explore individual differences during learning. Specifically, this study has explored how different individuals engage alternative learning subsystems (declarative vs. procedural).

To do so, the study has capitalized on the use of idiographic computational models, that is, 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; Collins, 2018; Daw, 2011).

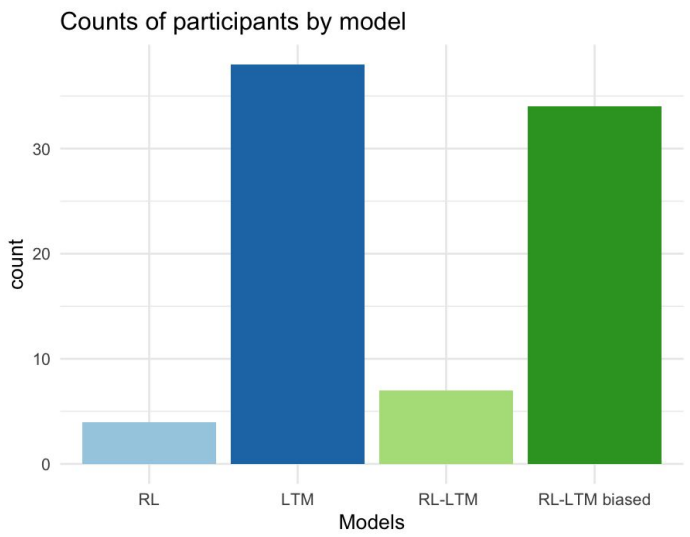

Figure 7: Count of the number of subjects that are best matched by each of the models.

It was found that different models fit different individuals, not only when models were effectively using different strategies (RL vs. declarative, LTM-based model) but also when one model was effectively nested within the other (basic models vs. integrated models). 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, with some adapting their learning strategy during the task, and some maintaining a bias towards one learning system. 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. 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. This will also afford us better fit between our models and behavioral data and parameter estimation than we have currently achieved.

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.

A second, point to be made concerns the importance of declarative memory in learning strategy, at least in humans, even in its long-term form. 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. Instead, our modeling results suggest that declarative-based models fit large sub-groups of individuals. Even the simplest, non-integrated model, accounts for 36% of our participants, and, altogether, models that at least include declarative components account for 73 out of 83 participants (Figure 7). Our results 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).

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 Set3 and Set4 conditions. Collins' (2018) explanation is that Set3 items are more likely to be still in working memory during learning, thus facilitating performance by direct reading of the associated response from a short-term buffer. Our 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 reasons, 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.

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).

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

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.

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.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.

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.