Dynamic Multi-level Learning in a Control-Learning Environment

Junwu Chen (junwu.chen@ugent.be)
Department of Experimental Psychology, Ghent University
Gent, B-9000, Belgium

Leslie Held (leslie.held@ugent.be)
Department of Experimental Psychology, Ghent University
Gent, B-9000, Belgium

Senne Braem (senne.braem@ugent.be)
Department of Experimental Psychology, Ghent University
Gent, B-9000, Belgium

Abstract

Humans are often confronted with multiple learning contingencies in real-life situations. Previous studies suggested that people tend to learn on different levels of abstraction sequentially rather than in parallel. In a reward learning environment where reward is only contingent on the task sequence (task switching/repetition), we employed a hierarchical reinforcement learning (HRL) model to investigate if individuals dynamically shifted the level at which they choose to learn over time. Our modelling analyses suggested that participants gradually shifted their priority from task-level learning to learning task sequences. Together, our findings are consistent with the sequential multi-learning hypothesis.

Keywords: Cognitive Control; Multi-level Learning; Hierarchical Reinforcement Learning.

Introduction

Humans can learn at multiple levels of abstraction (Abrahamse et al., 2016; Braem et al., 2024; Collins & Frank, 2013; Eckstein & Collins, 2020). However, in everyday life, multiple contingencies are available at different levels of abstraction, raising the question whether people learn about them serially or in parallel. Based on previous studies (Braem et al., 2024; Bugg, 2014; Held et al., 2024; Vallacher & Wegner, 1987), we argue that people may learn about different levels of abstraction in a more sequential manner rather than in parallel. More specifically, people may be potentially first biased towards attributing reinforcements to more concrete levels of information processing before turning to more abstract levels.

In the present study, as an initial exploration of this hypothesis, we developed a HRL model to examine across the whole learning process whether people will show a shift in prioritization among multiple levels of learning in a reward learning environment where reward is only contingent on performing task switching versus task repetition.

Task and Design

We employed a voluntary task-switching paradigm (Figure 1), in which participants can freely choose one of two tasks to perform in each trial. Specifically, participants were presented a word for each trial, and they can choose to judge either if the word is living or non-living (i.e., Animacy task) or whether it is larger than a basketball (i.e., Size task). 112 participants in total were recruited and randomly assigned to one of two reward environments (i.e., reward repeat/switch environment), where either task repetition or task switching was rewarded more points with a higher probability. Note that the probabilistic reward contingencies were exclusively determined by the task sequence, which means that the responses, as well as the tasks, shares almost equal reward contingencies. The reward probability varied across two experiments, and we combined them for all the following analyses (Exp1: 80%/20%; Exp2: 90%/10%).

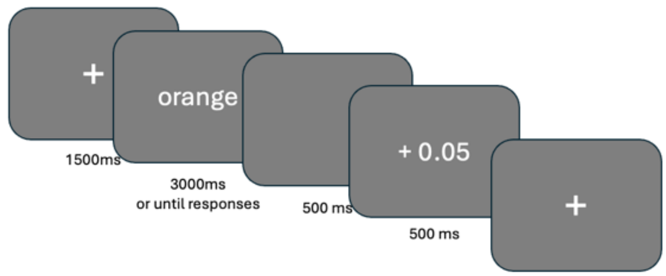

Figure 1. Task design.

Modeling Multi-level Learning

Our HRL model incorporates three levels of learning. Specifically, it simultaneously learns and updates the value of state-action pairs, tasks, and control over the task sequence (switch versus repeat). Critically, the learned values of different levels are integrated according to the following formulas, thus allowing all levels of learning to contribute to decision-making:

$$W_{int}(Task) = \varepsilon_{SI} \cdot Q(Task|Sequence) + (1 - \varepsilon_{SI}) \cdot Q(Task|Identity)$$

$$W_{int}(Response|State) = \varepsilon_{TR} \cdot W(Task|State) + (1 - \varepsilon_{TR}) \cdot Q(Response|State)$$

where the value of task sequence and task identity are integrated into the task-level weight, and this task-level weight is then integrated with the learned values of state-action pairs into the response-level weight. ε_{SI} and ε_{TR} are modelled as relative weighting factors with which these different levels are integrated. Ultimately, the integrated response-level weights are used to generate decisions using the SoftMax rule. The values learned at each level are independently updated using separate learning rates according to the delta learning rule.

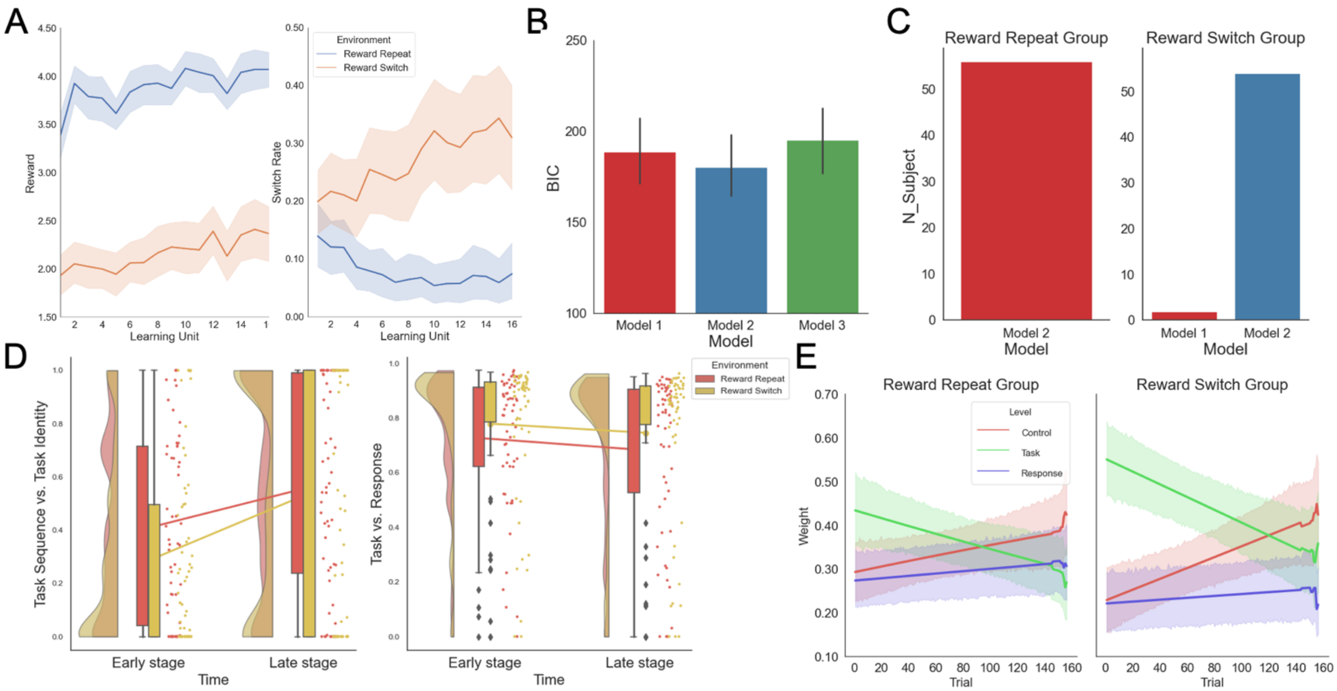

Figure 2. Behavioral and model fitting results. A) Learning curves of reward rate and switch rate. B) Group-level and C) individual-level model comparison results. D) Estimated weighting factors. E) Model predictions of how the relative weight of each level of learning varies across trials.

To investigate whether participants shifted their prioritized level of learning, we allowed the learning rates (i.e., Model 1), the integration weights (i.e., Model 2), or both (i.e., Model 3) to vary over time, which allowed us to investigate whether (and how) the learning rate and relative weight on each level evolved during the learning process.

Results and Discussion

Behavior To evaluate how reward and switch rates changed over time, we analyzed the data as a function of 16 time units (i.e., 10 trials per unit). As illustrated in Figure 2A, a linear mixed effect model showed a main effect of time on reward rate, $t = 3.76$, $p < .001$. Analyses on the switch rate indicated an interaction between group and time, $t = 6.67$, $p < .001$, showing that switch rate increased in the reward switch group, but decreased in the reward repeat group. These results suggested that participants gradually learned reward contingencies of the task sequence and adaptively configured more optimal levels of cognitive flexibility in both reward environments.

Model Fitting Group- and individual-level model comparisons both showed that Model 2 with the integration weights varying over time was the best-fitting model (Figure 2B-C).

As shown in Figure 2D, the weight ε_{SI} significantly increased across time, whereas another weight ε_{TR} significantly decreased as learning progressed. We recalculated the relative weights for each abstraction level and found that although the task-level weight was the largest at the very beginning, it decreased over time. In contrast, the control-level weight increased as learning progressed (Figure 2E). Collectively, our findings showed that as learning progressed, participants gradually shifted their prioritization from more concrete, subordinate task representations to more abstract, superordinate control representations.

Conclusion

Our results suggest that individuals dynamically shifted their prioritization of learning from first focusing on task-level learning to task-sequence-level learning. These findings provide preliminary evidence supporting the sequential multi-level learning hypothesis. Future directions are to develop and test a multi-level learning paradigm, and models that allow for different states of learning (rather than gradually evolving weights).

Acknowledgement

This project was supported by a BOF Starting grant and a FWO Grant (G0ABJ24N to S.B.).

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693–728. https://doi.org/10.1037/bul0000047

Braem, S., Chai, M., Held, L. K., & Xu, S. (2024). One cannot simply 'be flexible': Regulating control parameters requires learning. *Current Opinion in Behavioral Sciences*, *55*, 101347. https://doi.org/10.1016/j.cobeha.2023.101347

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587. https://doi.org/10.1037/a0035032

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852

Eckstein, M. K., & Collins, A. G. E. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, *117*(47), 29381–29389. https://doi.org/10.1073/pnas.1912330117

Held, L., Vermeylen, L., Dignath, D., Notebaert, W., Krebs, R., & Braem, S. (2024). Reinforcement learning of adaptive control strategies. *Communications Psychology*, *2*. https://doi.org/10.1038/s44271-024-00055-y

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3