Learning from free choices vs. learning from instructions - A student-teacher simulation study

Maayan Pereg,[1,2,3] Uri Hertz[4], and Nitzan Shahar[1,2]

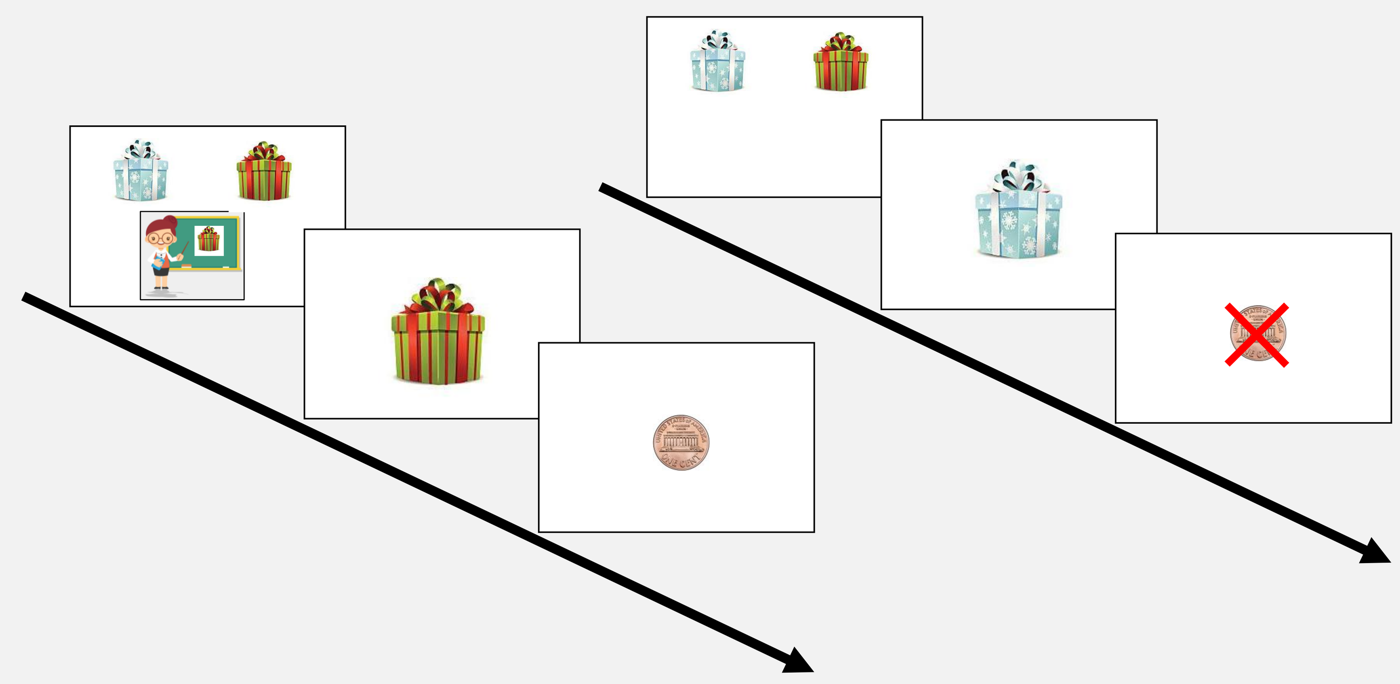
[1]School of Psychological Sciences, [2]Sagol school of Neuroscience, [3]Minducate science of learning research and innovation center, Tel-Aviv University, [4]Department of Cognitive Sciences, University of Haifa

INTRODUCTION

- While instructions-based actions could lead to better outcomes as they rely on an expert's knowledge (Pereg & Meiran, 2020), studies have shown superior learning from outcomes that follow free-choice actions (Cockburn et al., 2014; Ertmer et al., 2009).
- Cockburn et al., (2014) showed higher Dopamine bursts (supposedly reflecting learning) following free-choices relative to forced choices. However, individual differences related to this effect are yet unknown. Thus, teachers might face a challenge of whether to provide or withhold giving precise instructions (see Behrens et al., 2008).
- In this study we explored the student-teacher relationship using a **Reinforcement Learning** simulation study, where an Artificial Intelligence (AI) teacher is designed to learn the students' state, and select whether to provide or withhold giving instructions.

MODEL & SIMULATION

- Two-arm bandit task, with ten 50-trials blocks. True values were 0.35 and 0.65 for Bandit 1 and 2.
- Student learning was estimated using Q-learning (Rascorla-Wagner) as well as in a hybrid model involving associability-gated learning (Pearce-Hall, Li et al., 2011) .
- We describe three main free parameters: learning rate from instructions, free learning rate, and weight given to instructions (bias to dis/follow instructions).
- Proportion of instructions was determined by the AI teacher via Q-learning.



Student Q-learning model

$$PE_t = R - Q_{s(t)}$$

$$Q_{s(t+1)} = Q_{s_t} + \alpha_i PE_{(t)}$$

$$p = \frac{\exp(\beta * Q_{(t-1)})}{\sum \exp(\beta * Q_{(t-1)})}$$

Student hybrid model

$$Q_{s(t+1)} = Q_{s_t} + \alpha_i A_{(t)} PE_{(t)}$$

$$A_{s(t+1)} = w\left|PE_{(t)}\right| + (1-w)A_{s(t)}$$

Teacher Q-learning

$$Q_{inst(t+1)} = Q_{inst(t)} + \alpha_{teacher}((R * F_t) - Q_{inst(t)})$$

$$p_{inst} = \frac{\exp(\beta_{teacher} * Q_{inst})}{\sum \exp(\beta_{teacher} * Q_{inst})}$$

Teacher hybrid model

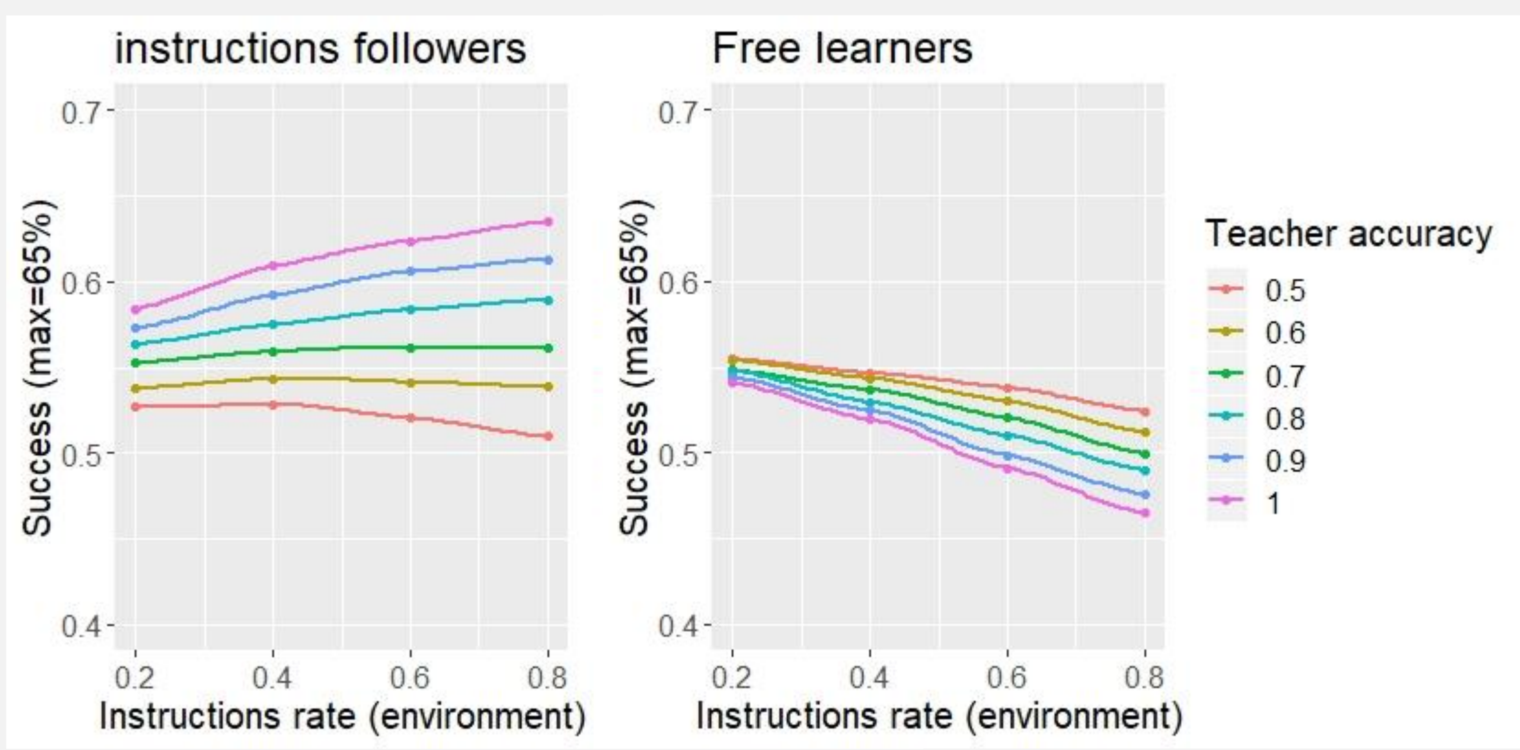
$$Q_{inst(t+1)} = Q_{inst(t)} + \alpha_{teacher} A_{(t)}((R * F_t) - Q_{inst(t)})$$

$$p_{inst} = \frac{\exp(\beta_{teacher} * Q_{inst})}{\sum \exp(\beta_{teacher} * Q_{inst})}$$

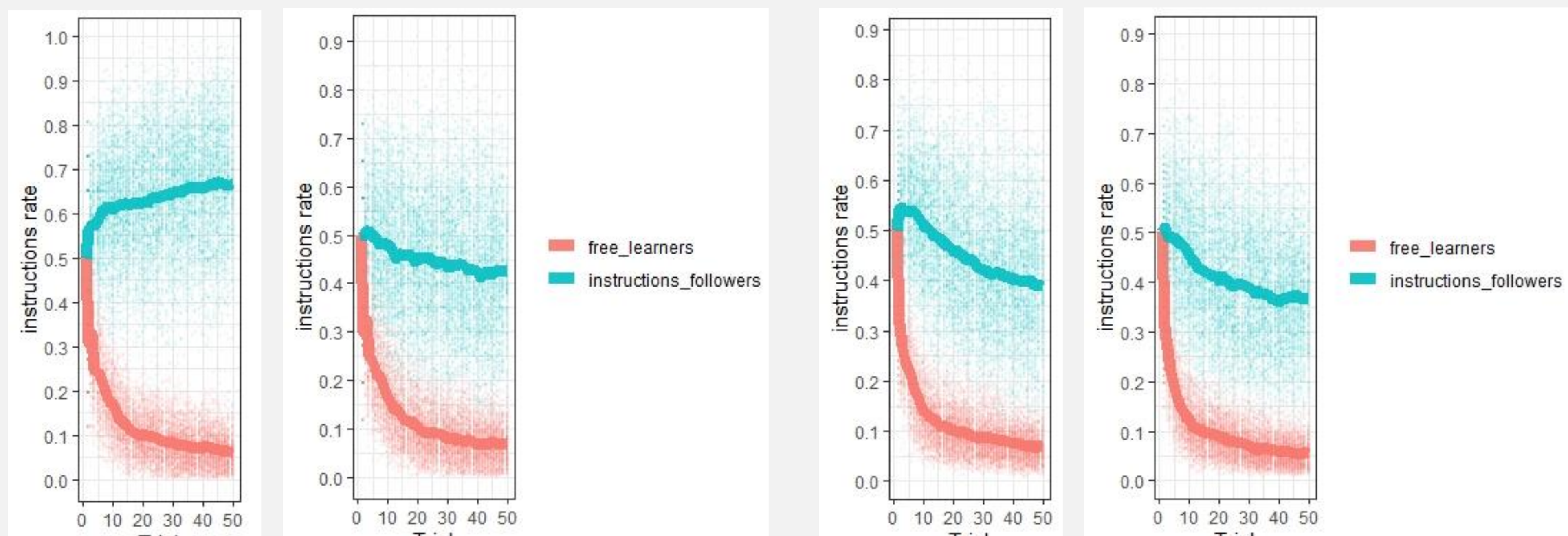
SIMULATION RESULTS

- Instructions-followers gain advantage when the environment involves high proportion of instructions, but only if the teacher is accurate.
- Free-learners gain advantage when the environment involves a low instructions-rate, regardless of teacher-accuracy.
- The AI teacher decreased instructions rate for free-learners, and increased it for instructions-followers. This trend was attenuated in the hybrid model.
- Free parameters recovery showed observed-expected correlations>0.90.

Estimated reward (task outcomes, hybrid model)



AI teacher adapted instructions rate throughout the task



DISCUSSION

- The suggested model is able to extract learning characteristics and show that all free parameters were interpretable.
- The in-silico results suggest that the rate of instructions is expected to influence learning outcomes for different agents.
- The study offers a step forward in the development and application of advanced personalized educational systems.

REFERENCES

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*, 245–249. doi:10.1038/nature07538

Cockburn, J., Collins, A. G., & Frank, M. J. (2014). A reinforcement learning mechanism responsible for the valuation of free choice. *Neuron, 83*(3), 551-557.

Ertmer, P. A., Glazewski, K. D., Jones, D., Ottenbreit-Leftwich, A., Goktas, Y., Collins, K., & Kocaman, A. (2009). Facilitating technology-enhanced problem-based learning (PBL) in the middle school classroom: An examination of how and why teachers adapt. *Journal of Interactive Learning Research, 20*(1), 35-54.

Pereg, M., & Meiran, N. (2020). Power of instructions for task implementation: superiority of explicitly instructed over inferred rules. *Psychological Research*, 1-19.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience, 14*(10), 1250-1252.

contact for correspondence: maayanru@gmail.com