Computational Modeling for Approach-Avoid Task with Reinforcement Learning Frameworks

Kai Hung[1], Eunice Yiu[2], & Alison Gopnik, Ph.D.[2,3]

1. Department of Computer Science, Rice University
2. Department of Psychology, University of California, Berkeley
3. Berkeley Artificial Intelligence Research, University of California, Berkeley

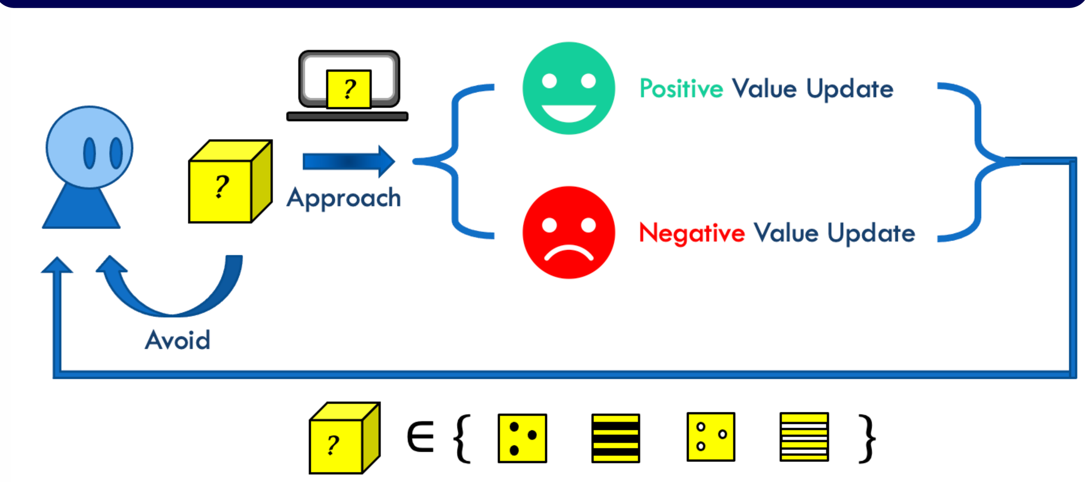
Abstract

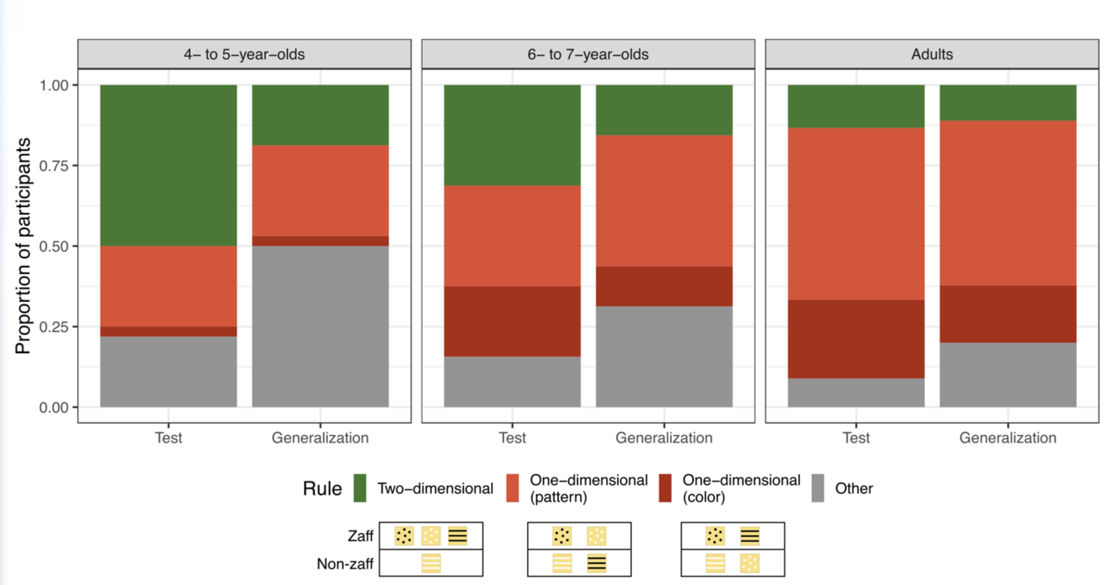
State-of-the-art machine learning algorithms have achieved incredible progress towards pattern recognition at the cost of large computing power and/or processed data, but their ability to generalize quickly and reliably remains poor relative to an average human child. To understand how children gather information and learn so much from so little, we focus on computationally modeling children's decision-making in an approach-avoid paradigm: children can opt to approach a certain stimulus, which may be rewarding or punishing; or they can opt to avoid it and learn nothing about whether the stimulus is rewarding or punishing. Specifically, we perform parameter estimation by fitting experimental data with variants of a standard reinforcement learning model including parameters such as learning rate and inverse temperature. Contrasting children's best-fit model parameters with adults, we find that children are more exploratory (lower inverse temperature) and less affected by external negative reward factors (smaller negative learning rate), yet potentially more capable of inferring the unusual conjunctive decision rule for maximizing net reward gains.

Background

- Developmental studies have shown that children are better learners than adults[1,2] and machines[3] in many situations. They tend to be more exploratory[1,2], resistant to various learning traps[1,4], and more able to reach correct conclusions via unusual conjunctive generalization rules[1,2].
- Increasingly, psychologists are drawing upon computational models[5], especially reinforcement learning paradigms[6, 7], to test specific hypothesis for human learning across development[8].
- A previous study showed that children are more exploratory than adults in an approach-avoid task and able to reach the correct conjunctive generalization rule more often[1]. However, it remains unclear how well their learning behavior may be captured by a reinforcement learning paradigm.

Experimental Design[1]





Model Formulation

Reinforcement Learning Framework (RL)[1]

Value Initialization $Q(a,s)_0 = \mathbb{E}(a)$

Decision Probability $P(a|s)_t = \frac{e^{\beta Q(a,s)_t}}{\sum_{a_i \in A} e^{\beta Q(a_i,s)_t}}$

Value Update $Q(a,s)_{t+1} = Q(a,s)_t + \alpha[r_t - Q(a,s)_t]$

RL Framework with +/- Learning Rate (RL2a)[1]

Value Update
$$Q(a,s)_{t+1} = \begin{cases} Q(a,s)_t + \alpha_+[r_t - Q(a,s)_t] & r_t \geq 0 \\ Q(a,s)_t + \alpha_-[r_t - Q(a,s)_t] & r_t < 0 \end{cases}$$

2-D RL Framework (RL-2D)

Dimension-Specific Value Functions $Q_{color}, Q_{pattern}$

Joint Value Function $Q = Q_{color} \times Q_{pattern}$

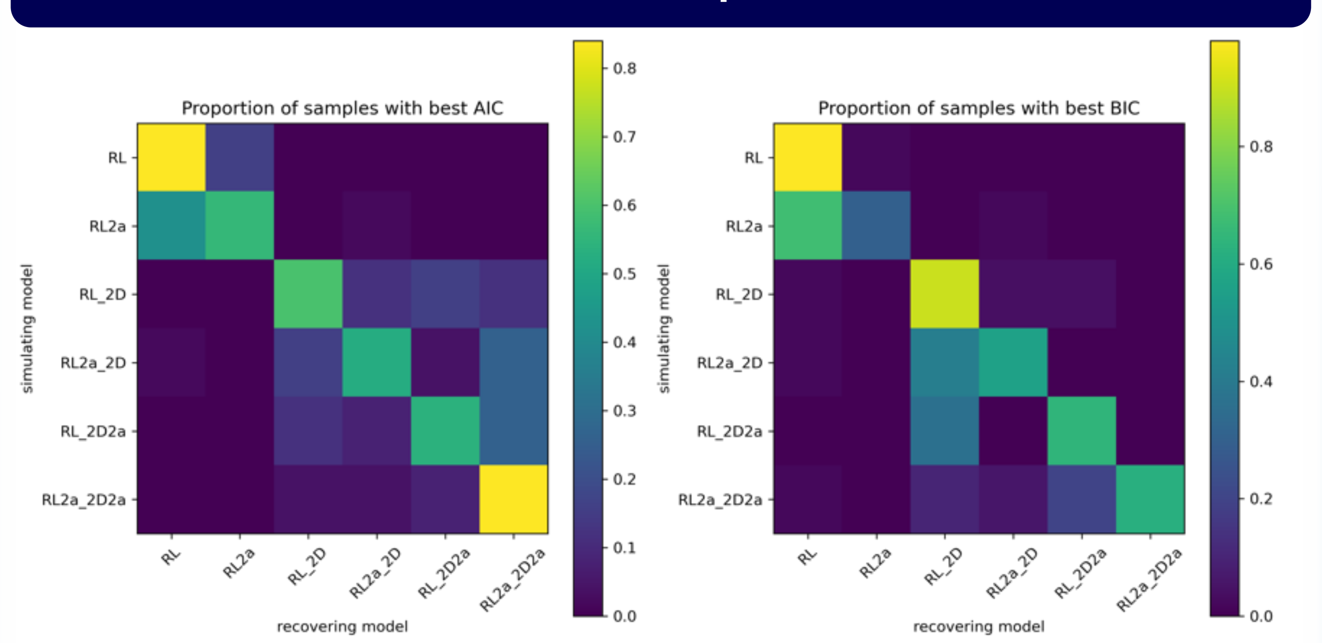
2-D RL Framework with Dimensional Learning Rates (RL-2D2a)

Dimension-Specific Value Updates

$$Q_{color}(a,s)_{t+1} = Q_{color}(a,s)_t + \alpha_{color}[r_t - Q_{color}(a,s)_t]$$

$$Q_{pattern}(a,s)_{t+1} = Q_{pattern}(a,s)_t + \alpha_{pattern}[r_t - Q_{color}(a,s)_t]$$

Model Comparison



AIC for Models across Age Groups

AIC	4-5 y/o's	6-7 y/o's	Adults
Baseline	709.78	709.78	1063.29
RL	583.72	464.93	561.46
RL2a	460.48	397.43	528.48
RL-2D	414.84	567.88	654.79
RL2a-2D	316.09	408.94	524.81
RL-2D2a	416.84	568.38	639.81
RL2a-2D2a	317.91	405.29	517.06

Model Performance

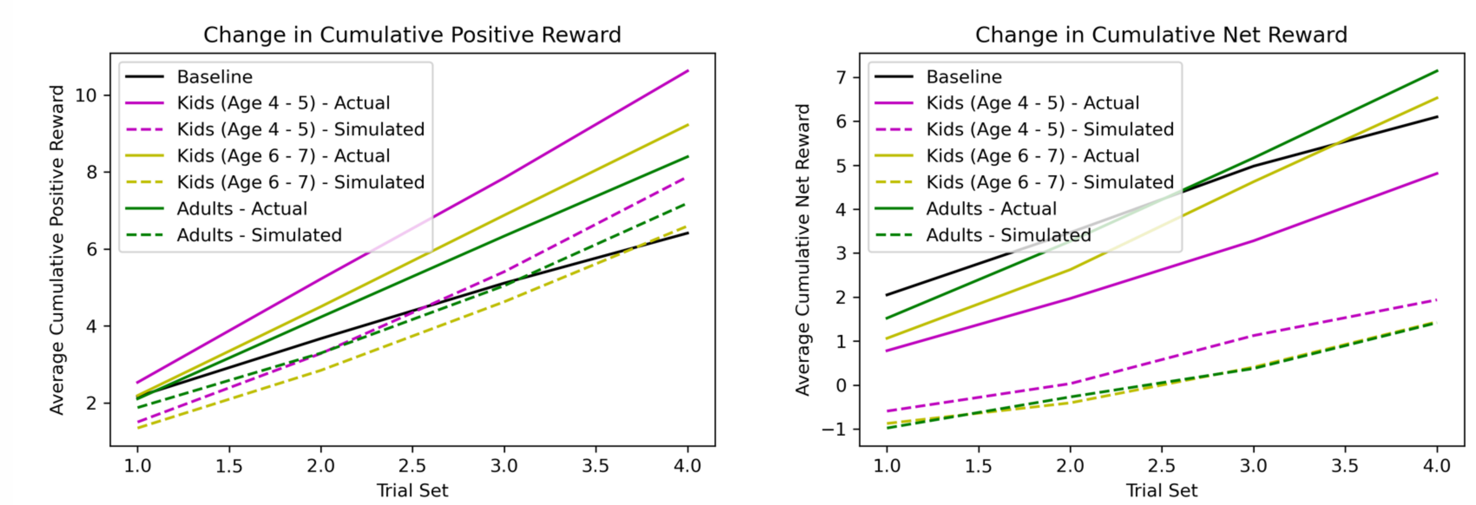
Best-Fit Model by Age Group (AIC)

4-5 years-old RL2a-2D
β: 3.497
α_+: 0.663
α_-: 0.01

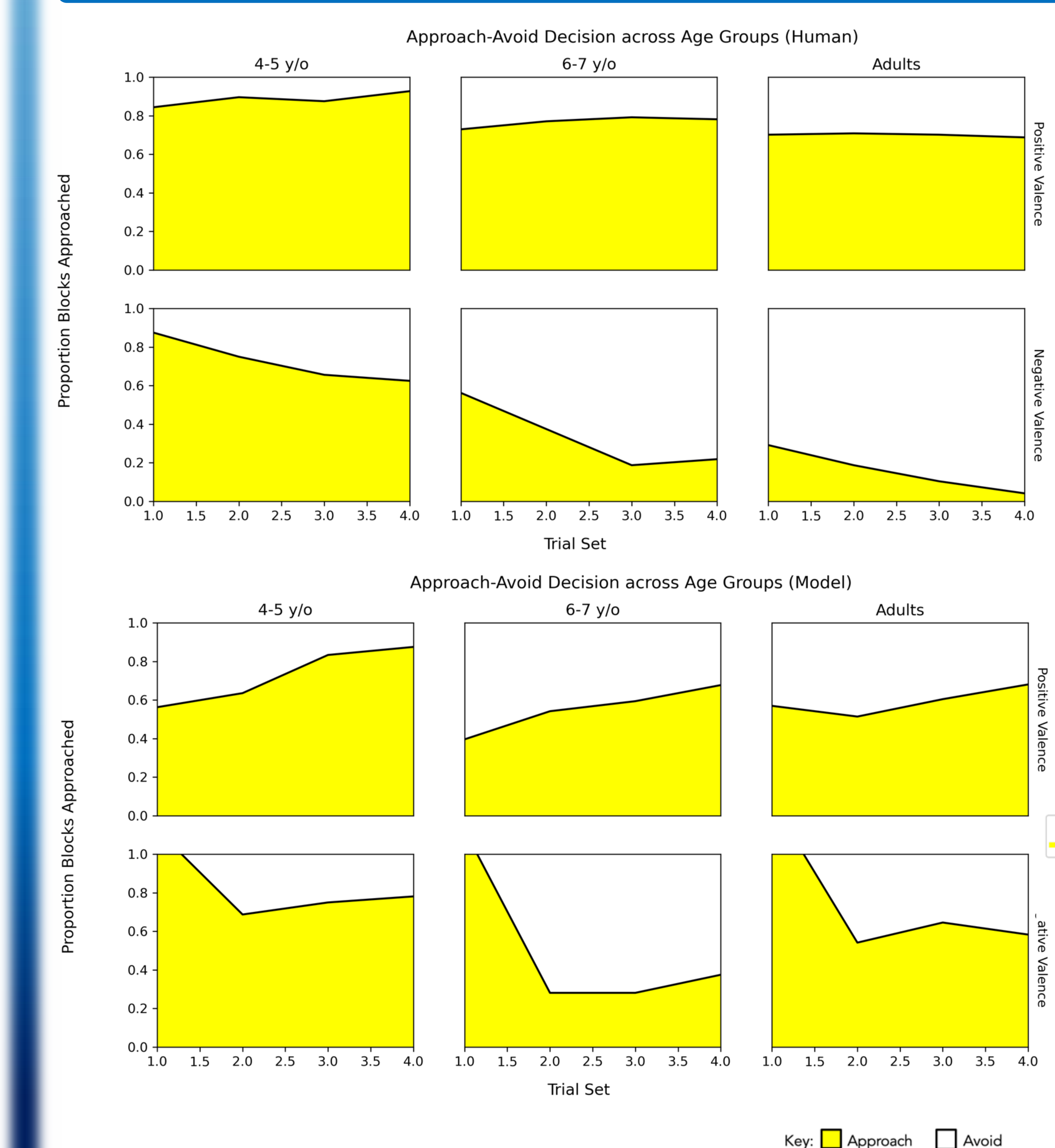
6-7 years-old RL2a
β: 2.223
α_+: 1.0
α_-: 0.01

Adults RL2a-2D2a
β: 5.437
$\alpha_{+,color}$: 0.572, $\alpha_{+,pattern}$: 0.428
$\alpha_{-,color}$: 0.043, $\alpha_{-,pattern}$: 0.124

Human vs. Model on Cumulative Reward



Human vs. Model Approach-Avoid Decisions



Conclusion

- Our best-fit computational models capture some exploratory trend of the participants, with the proportion of good valence blocks approached increasing as the we move onto older age groups.
- All the simulated agents exhibited greater rates of approaching non-valence blocks than their human counterparts across the trial time. This could be attributed to inverse temperature β's inability to model explore-exploit tradeoffs across time. This hypothesis also addresses the near zero best-fit negative learning rates α_- amongst children despite their demonstrated ability to avoid punishing blocks over time.
- The adult best-fit model contained both value function for color and pattern. This is consistent with their tendency to generalize to a one-dimensional rule. Furthermore, $\alpha_{-,pattern} > \alpha_{-,color}$ suggests greater sensitivity towards punishment with respect to block pattern, which is consistent with adults' tendency to infer a 1-D color rule.

Future Works

- Introduce a "curiosity" factor to model the directed exploration of the children[9].
- Introduce a free parameter to measure the extent of an agent's generalization towards a 1-D rule as opposed to a 2-D one.
- Adapt a Bayesian paradigm to introduce prior beliefs about color, pattern, etc.

References

1. Liquin, E. & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition, 218.* https://doi.org/10.1016/j.cognition.2021.104940
2. Lucas, C., Bridgers, S., et al. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition, 131(2)*, 284-299. https://doi.org/10.1016/j.cognition.2013.12.010
3. Kosoy, E., Collins, J., et al. (2020). Exploring Exploration: Comparing Children with RL Agents in Unified Environments. *arXiv:2005.02880.* https://doi.org/10.48550/arXiv.2005.02880
4. Rich, A. S., Gureckis, T. M. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General, 147*(11). 1553-1570. https://doi.org/10.1037/xge0000466
5. Wilson, R. C., Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife, 8.* https://doi.org/10.7554/eLife.49547
6. Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology, 53*(3), 139-154. https://doi.org/10.1016/j.jmp.2008.12.005
7. Sutton, R. S. & Barto, A. G. (2018). *Reinforcement Learning: An Introduction.* MIT Press.
8. Nussenbaum, K. & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience, 40.* https://doi.org/10.1016/j.dcn.2019.100733
9. Pathak, D., Agrawal, P., et al. (2017). Curiosity-driven Exploration by Self-supervised Prediction. *arXiv:1705.05363.* https://doi.org/10.48550/arXiv.1705.05363

Acknowledgement

This work was under the supervision of Eunice Yiu. Funding provided through the NSF SUPER REU program at University of California, Berkeley. Idea for the joint value function in 2-D RL belongs to Fei Dai,

For more information, you can reach me at kai.hung@rice.edu and view my LinkedIn profile via the QR code.