Computational Modeling for Approach-Avoid Task with Reinforcement Learning Frameworks

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

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

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

THE UNIVERSITY OF CALIFORNIA BERKELEY · LET THERE BE LIGHT · 1868

Supported by NSF SUPERB REU Program

Outline

- Motivation
- Previous Work & Experiment Set-Up
- Our Approach: What, Why, and How
- Model Formulation
- Model Comparison
- Results
- Conclusion
- Future Works
- Acknowledgements

Motivation

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?

Alan Turing, 1950.

Motivation

Modern AI Frameworks

Photo Credit: Shutterstock

- ➢ Ex. supervised, reinforcement learning
- ➢ Need lots of data
- ➢ Not much (or right) generalization
- ➢ Pattern recognition, no account for causality

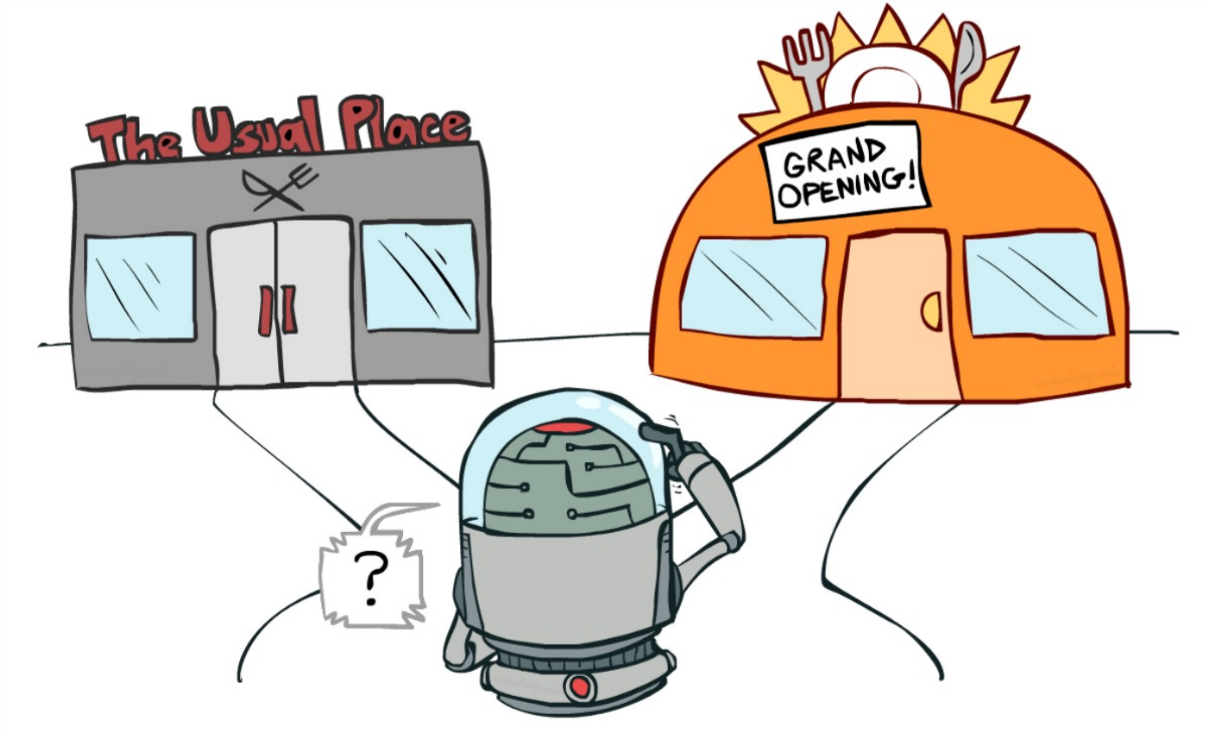
4-Year-Olds

Photo Credit: Raising Children Network

- ➢ Little supervision or reinforcement
- ➢ Very little data
- ➢ Excellent generalization
- ➢ Ability to form causal predictions

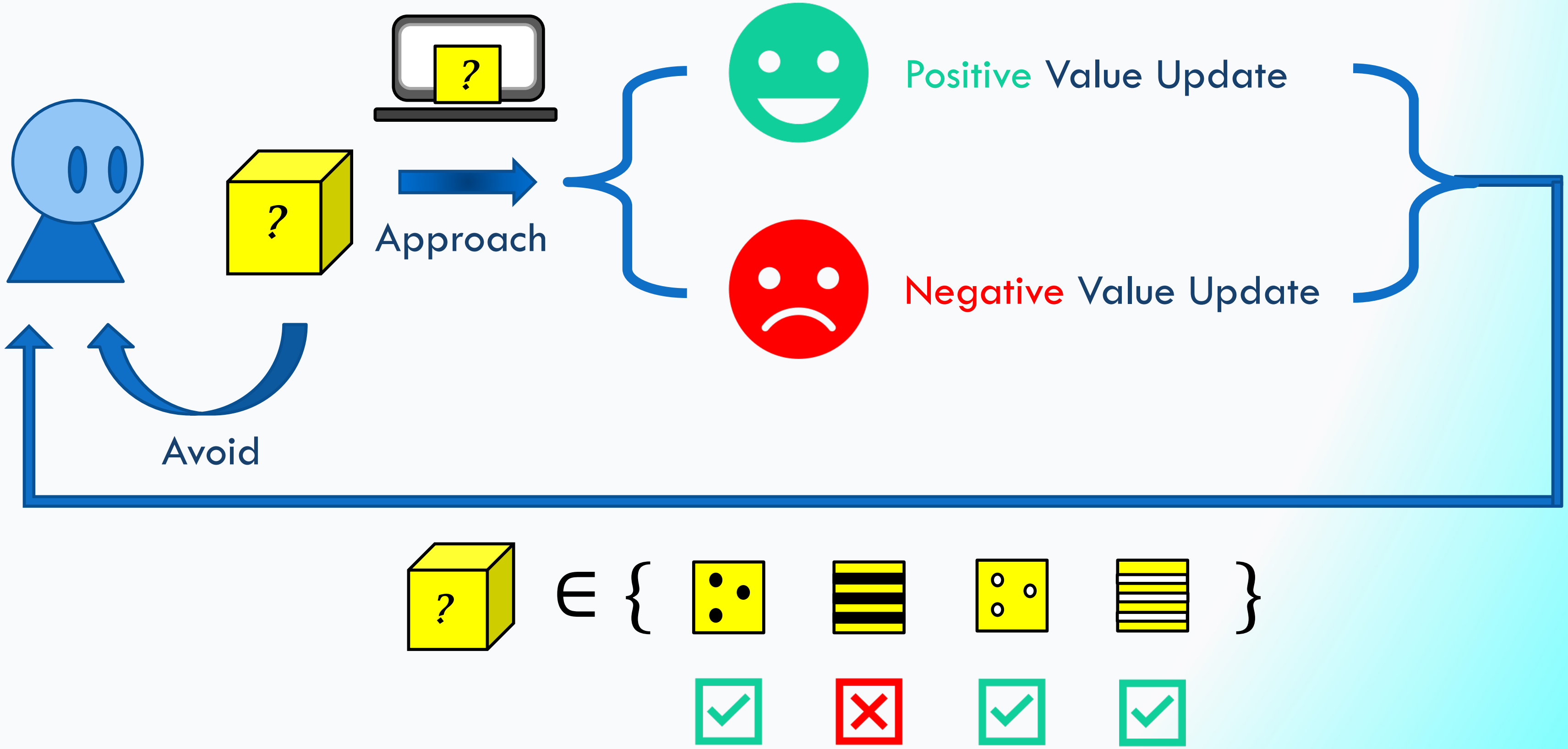
Motivation

Explore-Exploit Tradeoff



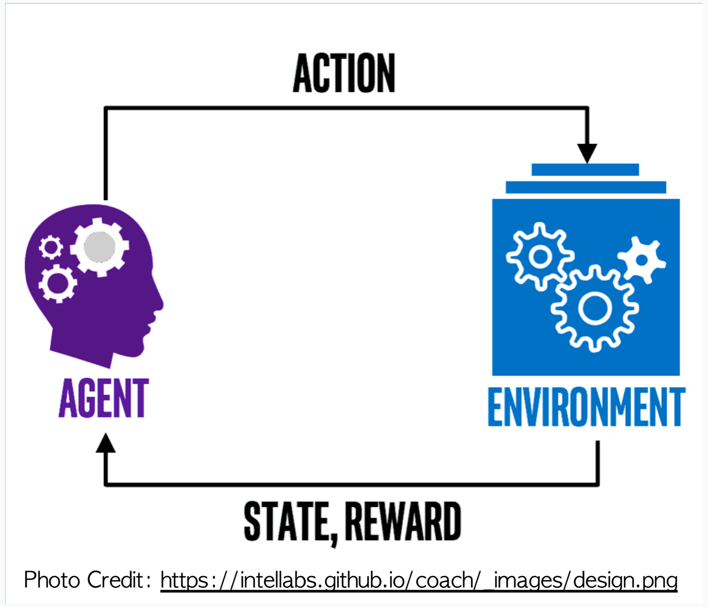
Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philosophical Transactions B, 375.* https://doi.org/10.1098/rstb.2019.0502

Previous Study

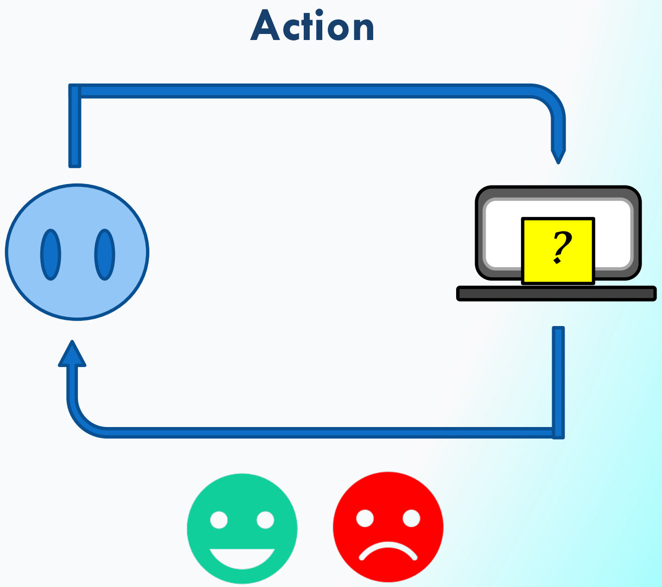


Liquin, E. & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition, 218.* https://doi.org/10.1016/j.cognition.2021.104940

Our Approach: What, Why, and How



Reinforcement Learning

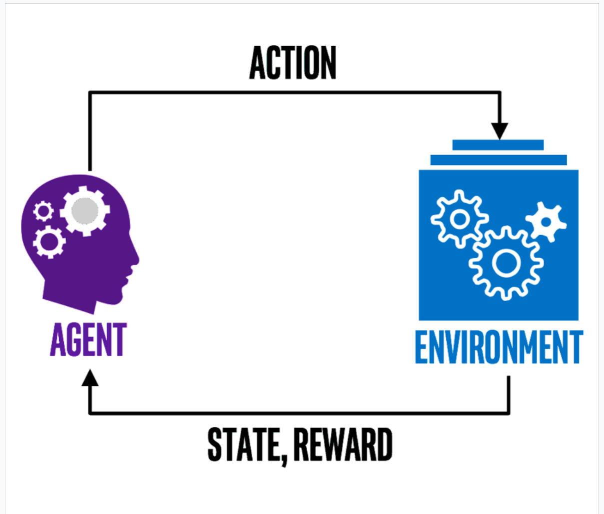


Experiment Design

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

Reinforcement Learning (RL) Model

Definition: Q-Learning



Value Update Mechanism	$Q(a,s)_{t+1} = Q(a,s)_t + \alpha[r_t - Q(a,s)_t]$
Decision Probability	$P(a\|s)_t = \frac{e^{\beta Q(a,s)_t}}{\sum_{a_i \in A} e^{\beta Q(a_i,s)_t}}$
Parameters of Interest	Learning Rate α, Inverse Temperature β

Photo Credit: https://intellabs.github.io/coach/_images/design.png

Reinforcement Learning (RL) Model

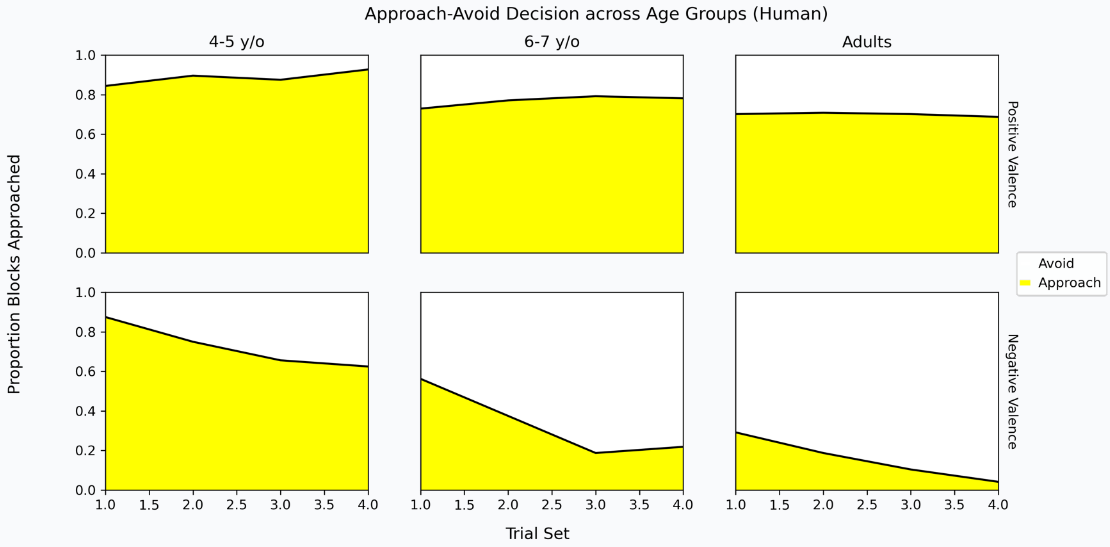
Parameter Estimation

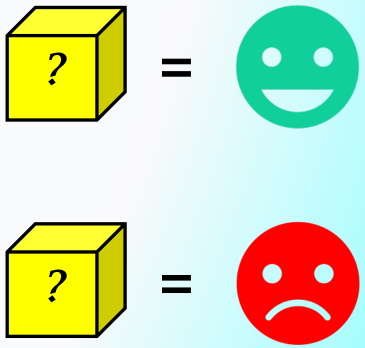
Parameter Estimation $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \boldsymbol{P}(\boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{M})$

4-5 years-old	6-7 years-old	Adults
$\hat{\alpha}$: 1.0	$\hat{\alpha}$: 1.0	$\hat{\alpha}$: 0.819
$\hat{\beta}$: 0.536	$\hat{\beta}$: 1.364	$\hat{\beta}$: 2.369

Advanced RL Models

RL2a: Positive & Negative Learning Rates a_+, a_-





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

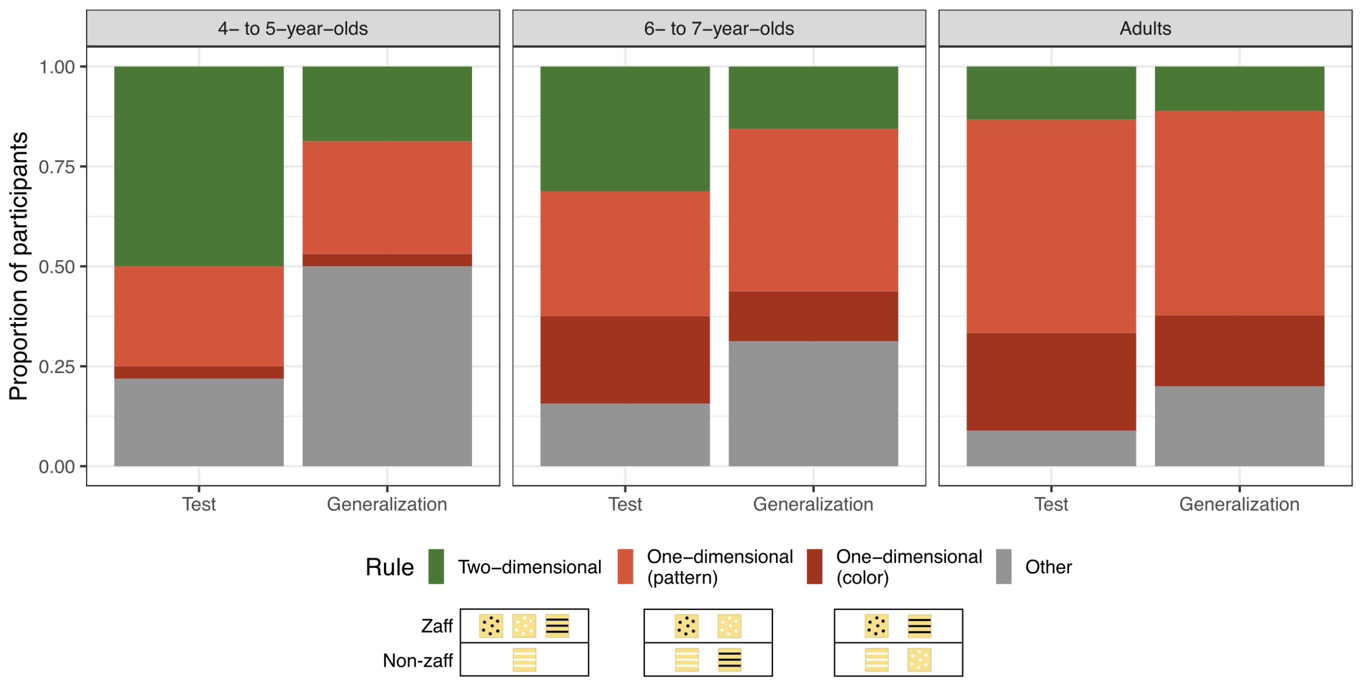
Cazé, R.D., van der Meer, M.A.A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biol Cybern*, *107*, 711-719. https://doi.org/10.1007/s00422-013-0571-5

Advanced RL Models

RL-2D: Dimension-based Value Functions

Color-Pattern Value Functions $Q_{color}, Q_{pattern}$

Joint Value Function $Q(a, s) = Q_{color}(a, s) \times Q_{pattern}(a, s)$[1]



Credit to Fei Dai (University of California, San Diego) for idea towards joining the two value functions.

Advanced RL Models

RL-2D: Dimension-based Value Functions

Color-Pattern Value Functions $Q_{color}, Q_{pattern}$

Joint Value Function $Q(a,s) = Q_{color}(a,s) \times Q_{pattern}(a,s)$[1]

RL-2D2a: 2-D with Dimension Learning Rates $a_{color}, a_{pattern}$

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

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

Credit to Fei Dai (University of California, San Diego) for idea towards joining the two value functions.

Model Comparison

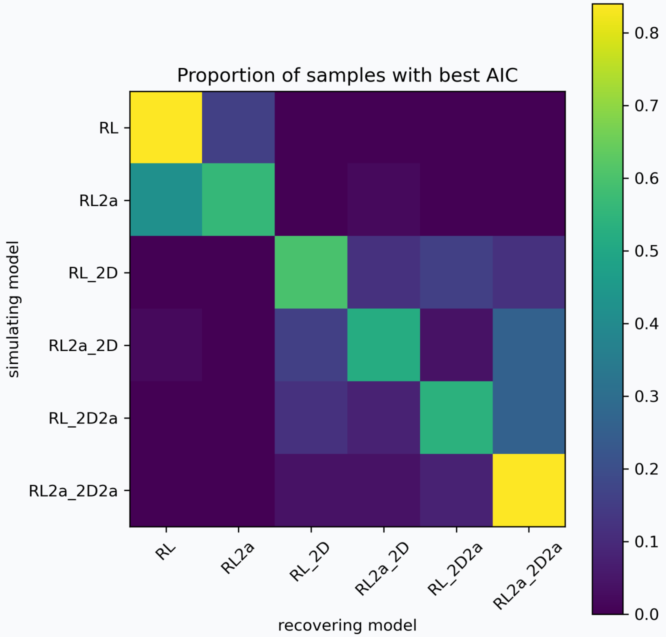
Akaike Information Criterion

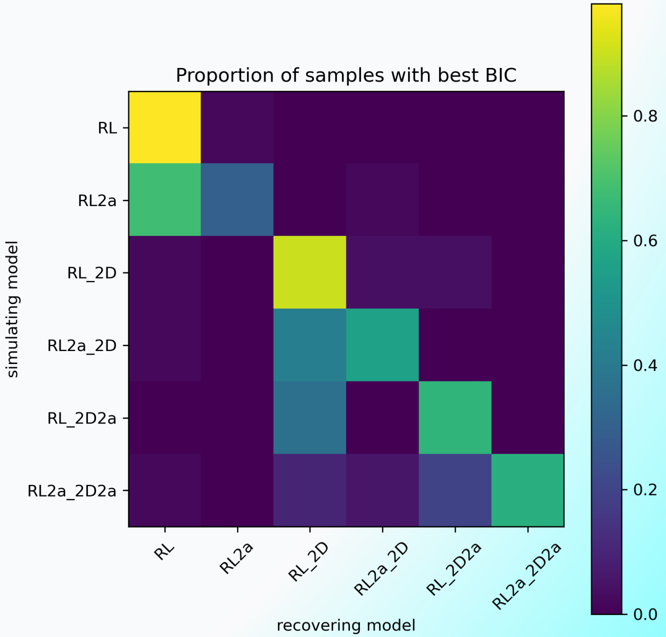
$$AIC = 2k \; - 2\ln(\hat{L})$$

Bayesian Information Criterion

$$BIC = k\ln(n) \; - 2\ln(\hat{L})$$

k = number of estimated parameters in the model
$\hat{L}$ = maximum value of the likelihood function for the model
n = number of observations





Model Comparison: Best Models

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 Comparison: Best Models

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 Comparison: Best Models

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

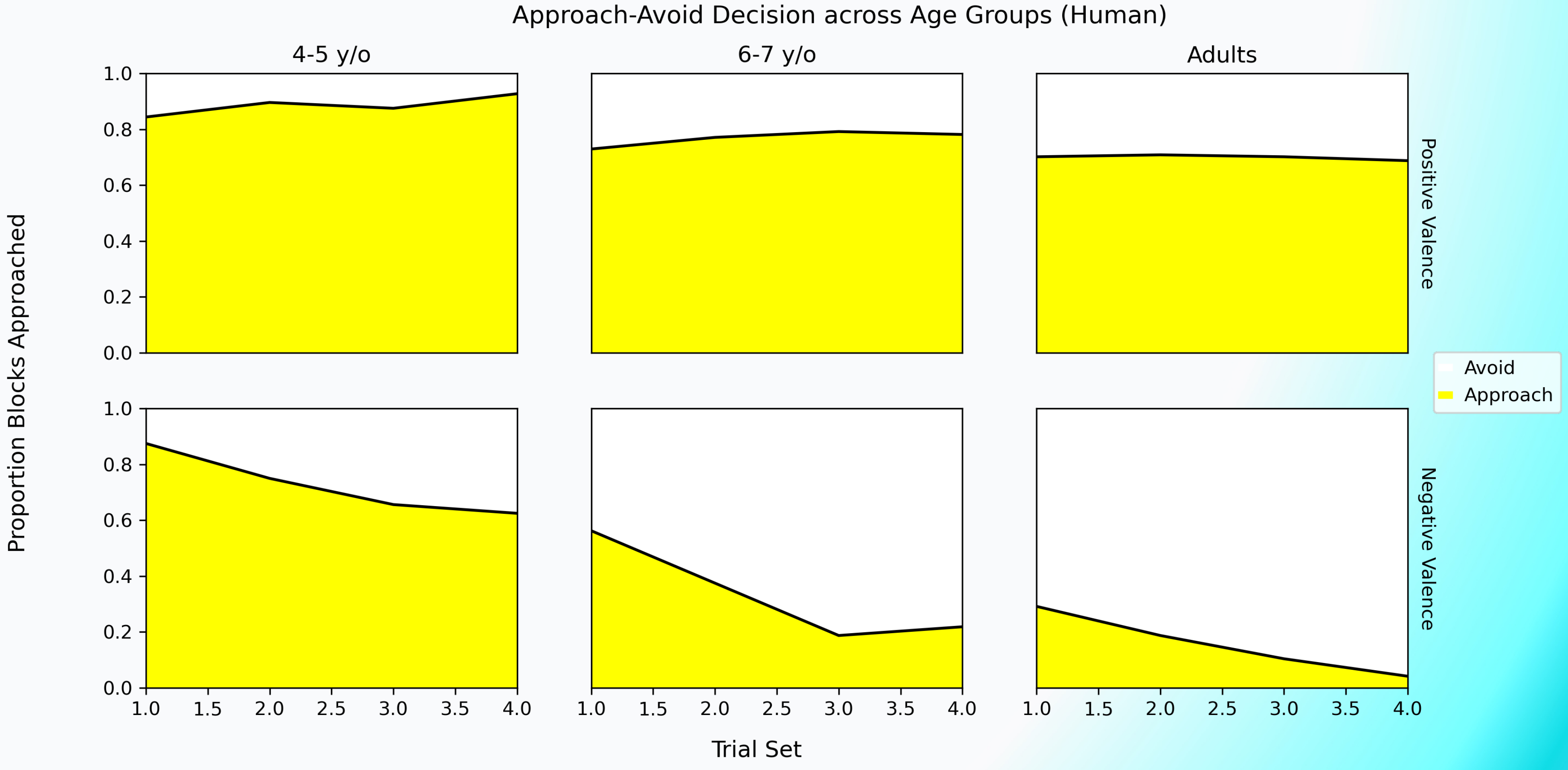
Adults

RL2a-2D2a

- β: 5.437
- $\alpha_{+,\mathrm{color}}$: 0.572
- $\alpha_{-,\mathrm{color}}$: 0.043
- $\alpha_{+,pattern}$: 0.428
- $\alpha_{-\ pattern}$: 0.124

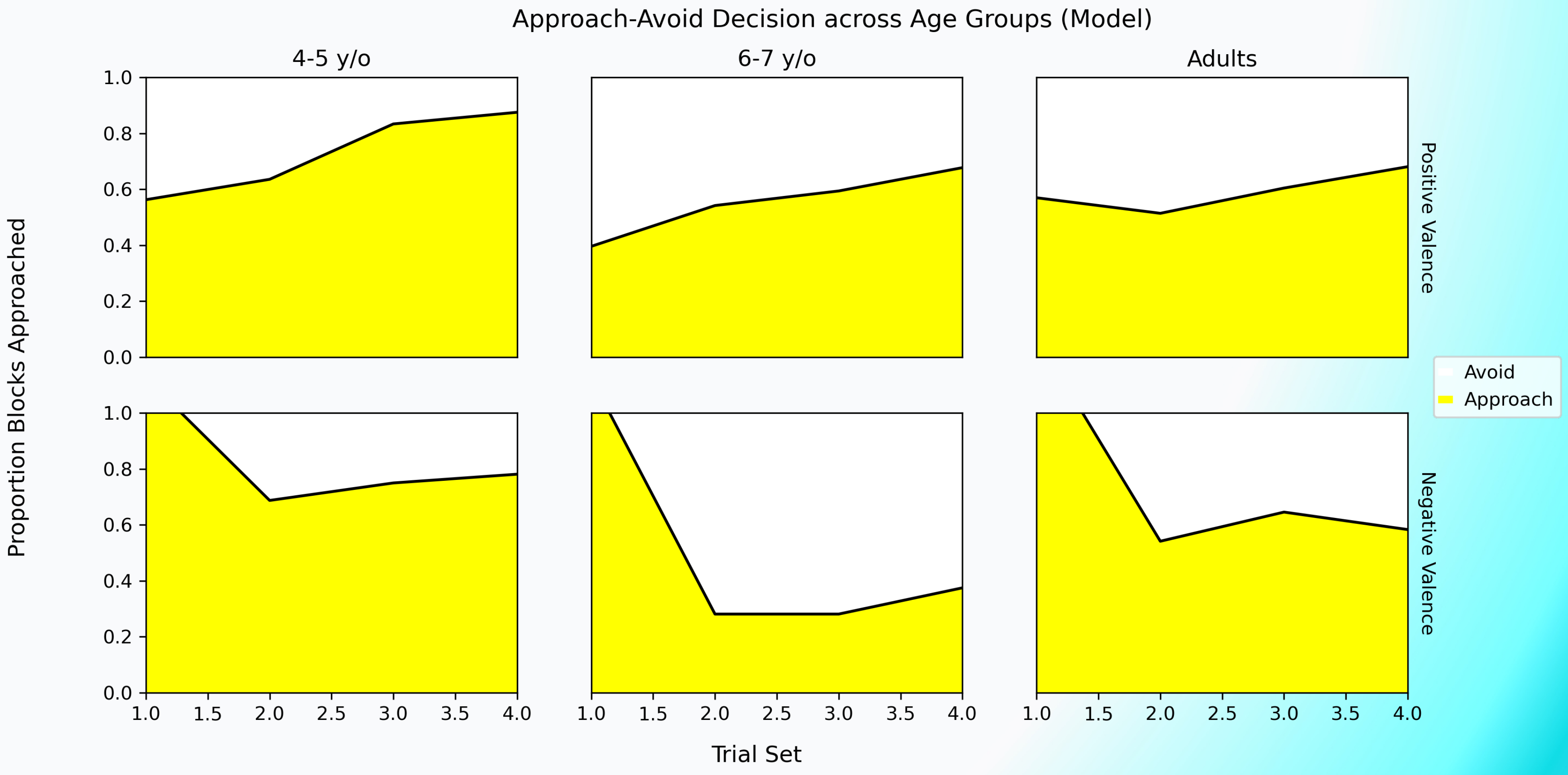
Model Performance vs. Human

Proportion of Approach-Avoid (Humans)



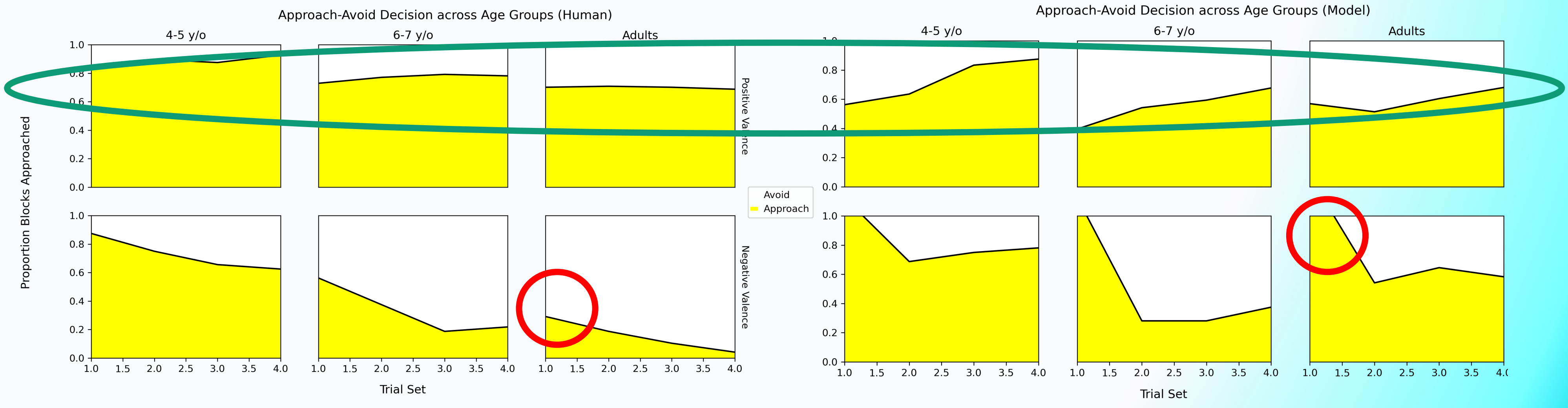
Model Performance vs. Human

Proportion of Approach-Avoid (Models)



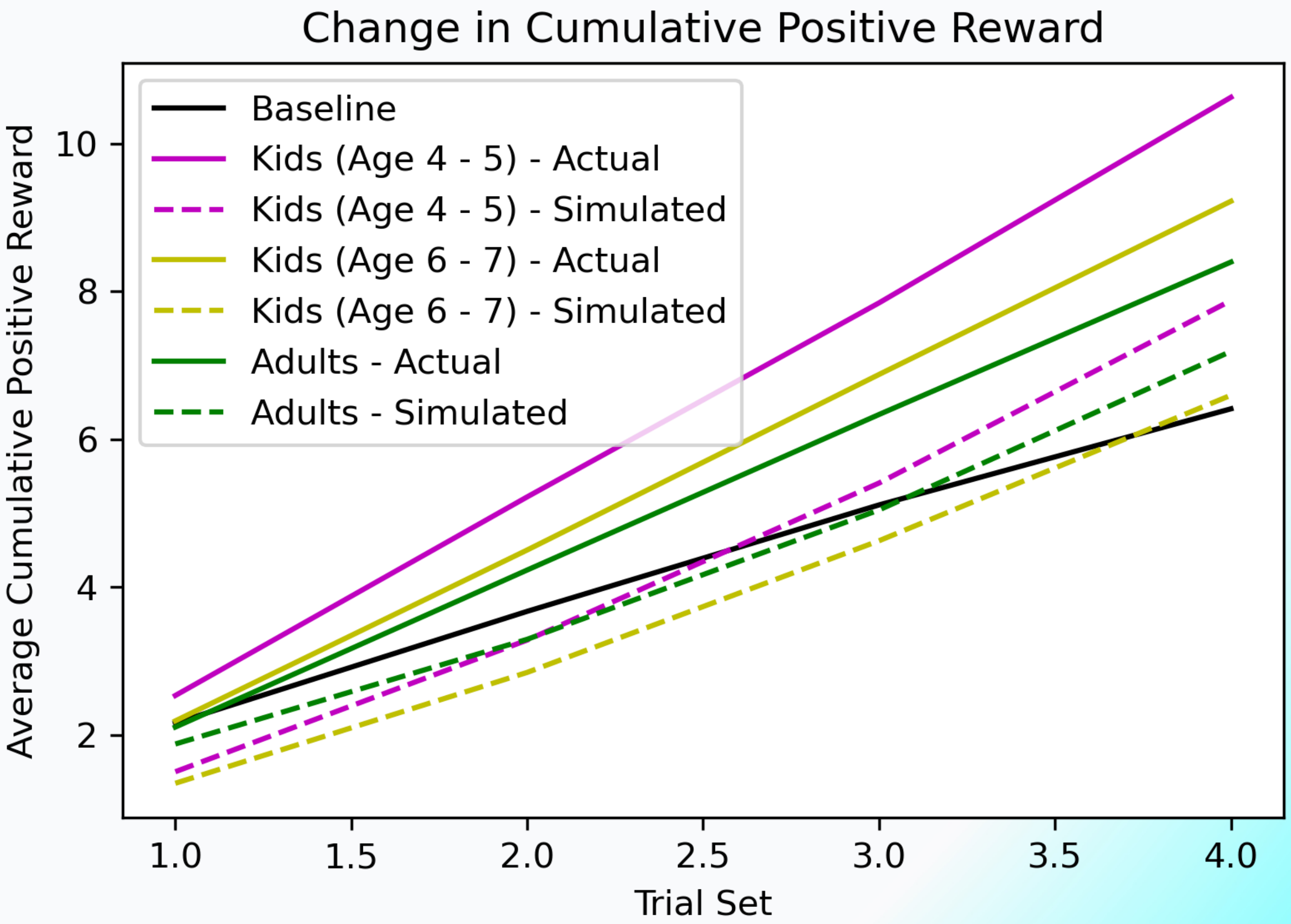
Model Performance vs. Human

Proportion of Approach-Avoid (Humans)



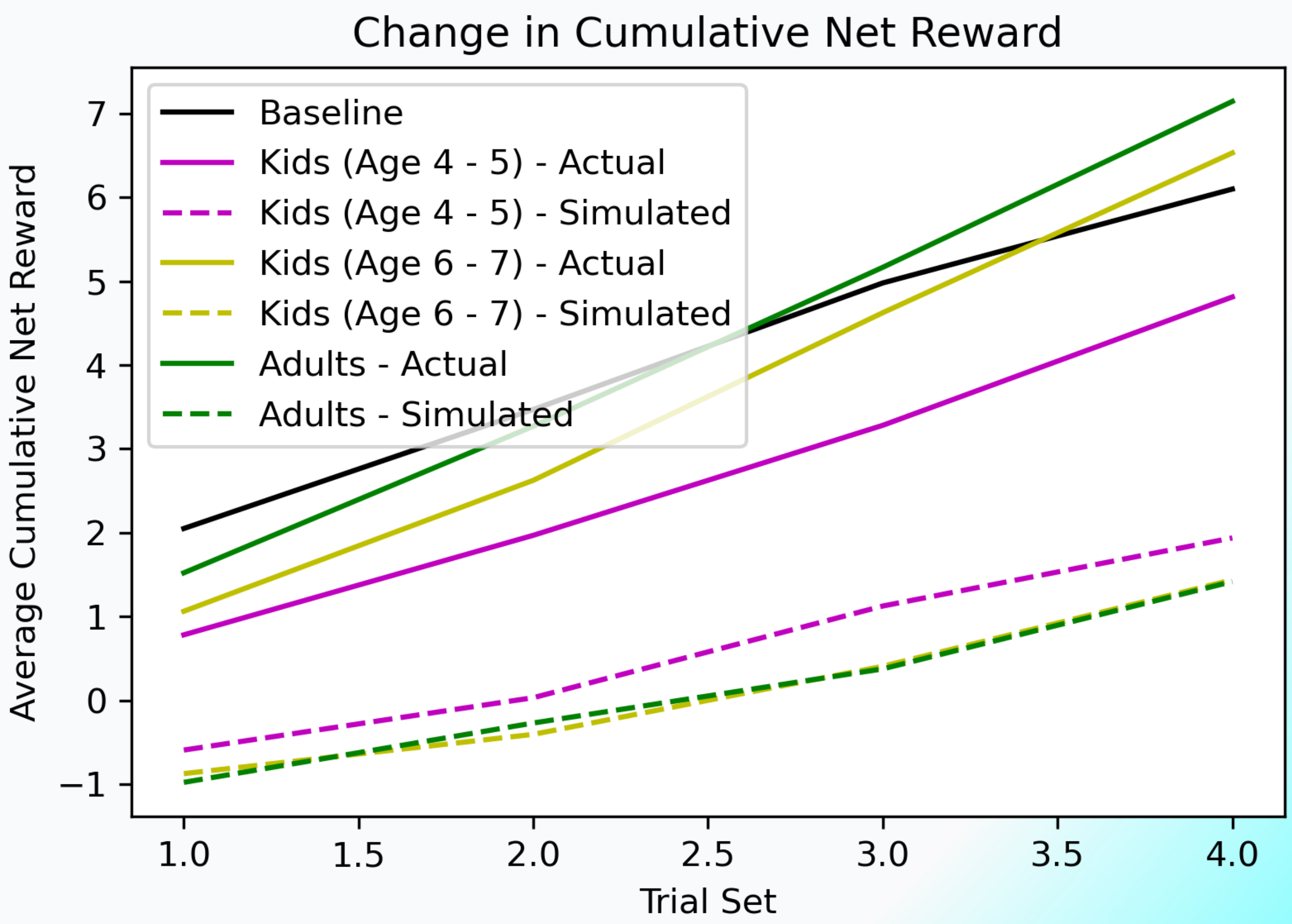
Model Performance vs. Human

Change in Cumulative Positive Reward



Model Performance vs. Human

Change in Cumulative Net Reward

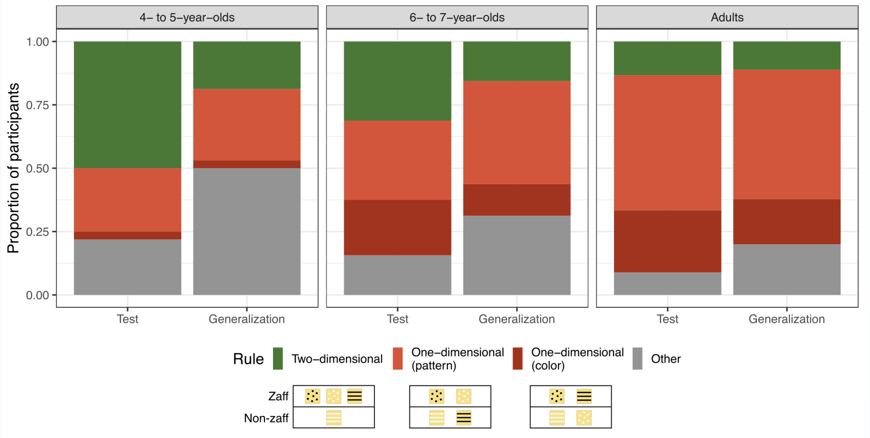


Results

Best-Fit Model for Adult (RL2a-2D2a)

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

$\alpha_{-,pattern} > \alpha_{-,color}$ suggests that the participants are more sensitive to negative reward associated with the pattern than color.



A sensitivity to negative stimuli on pattern is consistent with how more adults conform to a one-dimensional pattern rule since early generalization means they will grow avoidant to objects based on their pattern.

Conclusion & Future Works

- Despite popular comparisons between reinforcement learning and human learning, our models struggle to replicate the behavior of their human counterparts particularly in terms of negative stimulus.

- As a future direction, we will consider components that capture **curiosity** or **directed exploration.** It appears that the more exploratory human participants are conducting a strategic search to obtain information, which cannot be captured by our inverse temperature β parameter.

- We may also explore the use of Bayesian paradigms rather than RL paradigms, which allows us to consider the reinforcement process as one of updating prior beliefs.

Acknowledgements

Thank you to our program director, Leslie Mach[1], for coordinating and the National Science Foundation for funding this research opportunity at the University of California, Berkeley.

Thank you to Fei Dai[2, 3], David Chan[1], and Milena Rmus[2] for suggestions and help in refining the computational models during the early stages of this project.

I would also like to thank Dr. Alison Gopnik[2], Rose Reagan[2], Dr. Benjamin Pitt[2], other members of the Gopnik Cognitive Development & Learning Lab[2], and all the graduate student mentors from BAIR (Brent, Leyla, Ruchir, Rudy) for a welcoming and fun research environment.

Lastly, the project and my summer could not be this fruitful without the continuous mentorship and support of my mentor, Eunice Yiu[2]. My time would also be incomplete without the friendship of my fellow SUPERB cohort.

1. Department of Electrical Engineering and Computer Science, University of California, Berkeley
2. Department of Psychology, University of California, Berkeley
3. Department of Computer Science, University of California, San Diego

Questions?

Happy to discuss more during the poster session or over email!

Email: kai.hung@rice.edu