Exploring NeuroAI Models of How Learned Behavior Can Evolve Into Instinct

Christos Karaneen, Arkarup Banerjee

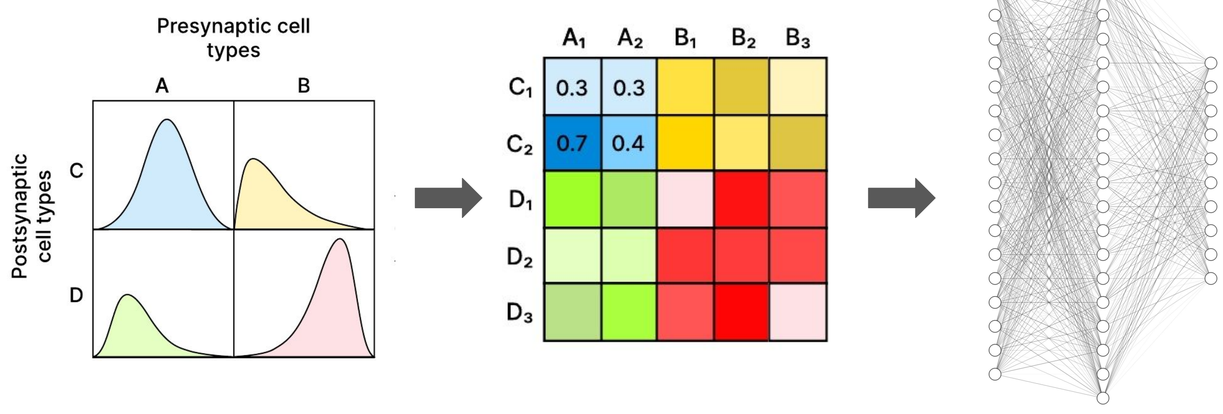
Cold Spring Harbor Laboratory

Abstract

- How do **instinctual** behaviors **evolve?**
- The **Baldwin effect** describes how *learned* behaviors can evolve into more *innate* ones
- We show that selection based on **learning** ability decreases the time required to achieve task performance threshold compared to selection based on **innate** ability

Introduction

- We construct artificial agents whose genotypes encode neural net **parameters** and **learning rates**



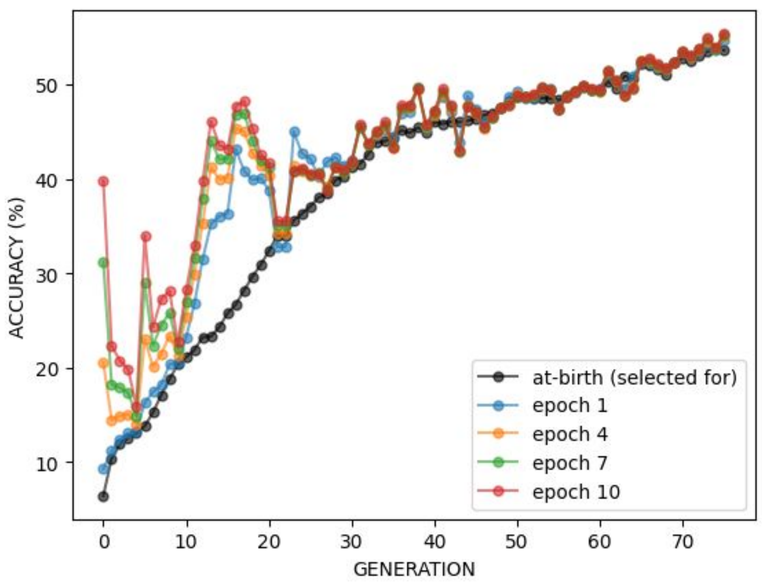
Lachi et al., 2024

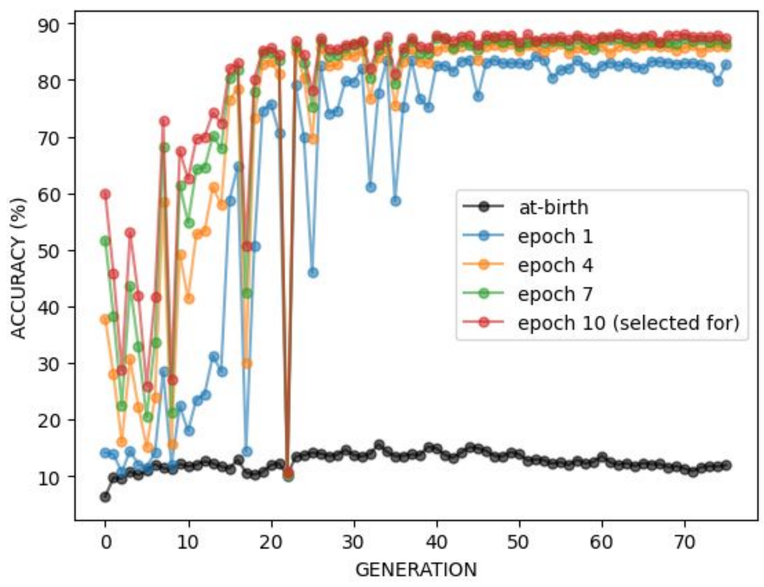
- $\mathbf{w}_{ij} \sim G_w = N(\mu, \sigma^2)_{ij}$ and $\mathbf{lr} \sim G_{lr} = \text{Lognormal}(\mu, \sigma^2)$
- Evolution simulation across generations:
  - Select best agents from population and produce new one:

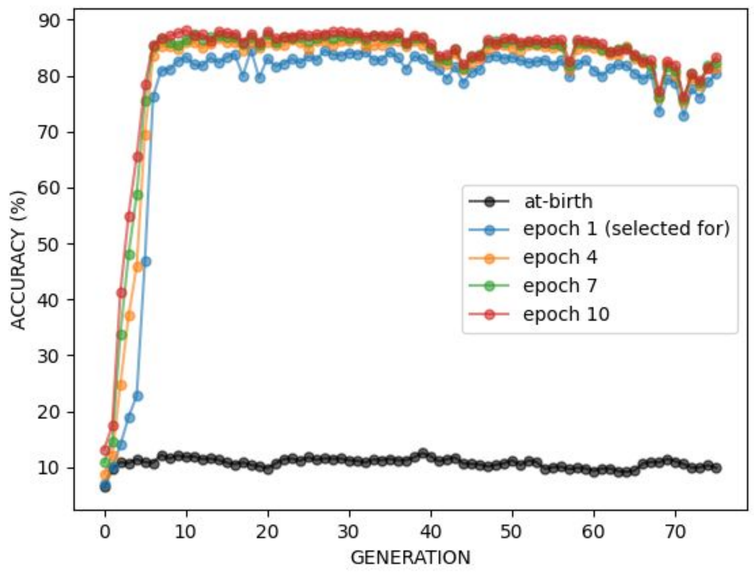
$$p_{new} = \alpha * p_1 + (1 - \alpha) * p_2$$

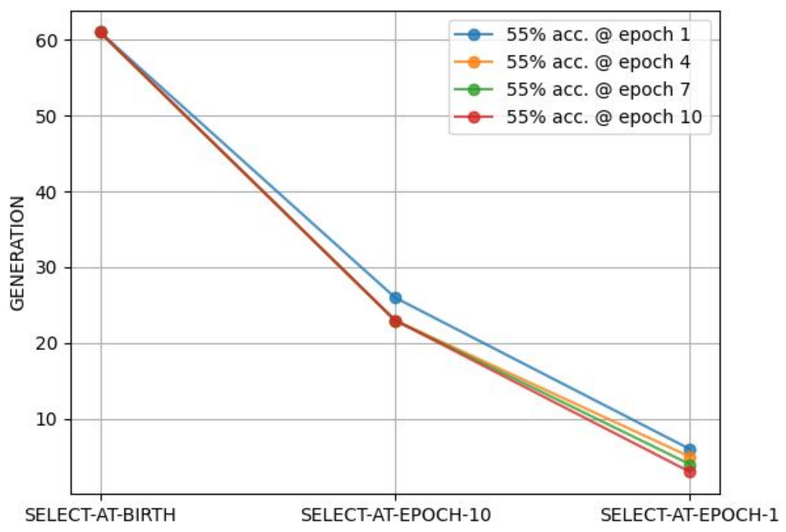
  - Train neural net for 10 epochs per generation

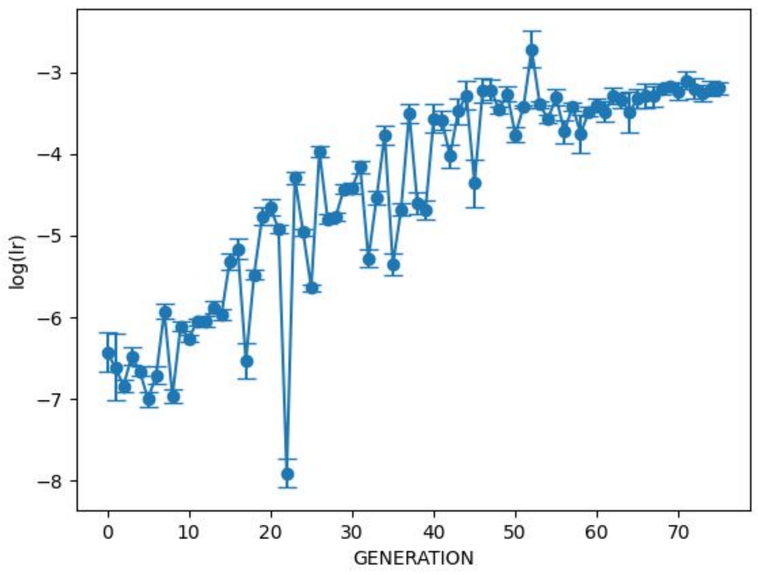
Results

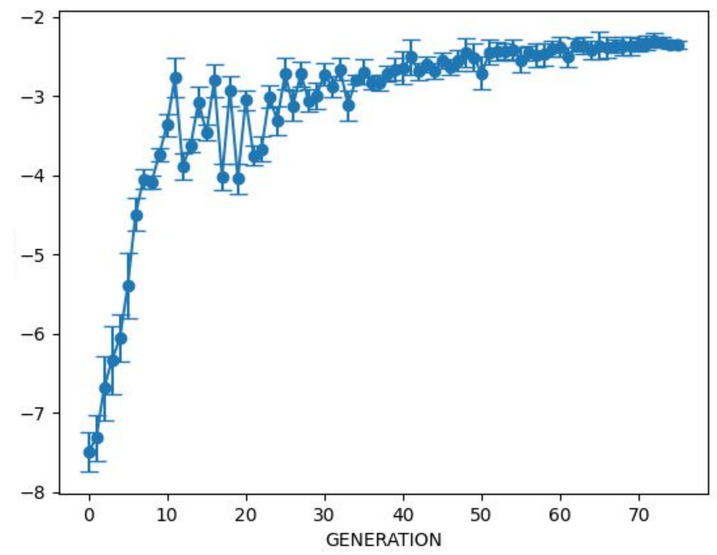












Conclusions

- Learning accelerates evolution:
  - Selecting on performance **after-training** decreases time to saturation compared to selection **at-birth**
  - Selecting **earlier** in the learning trajectory leads to faster saturation than selecting **later** does
- A first step towards modeling how learned behaviors can evolve into more innate ones

Acknowledgements

This work was supported by **Cold Spring Harbor Laboratory's *NeuroAI* program.**

Contact information

ckarageorgkaneen@gmail.com

github.com/ckarageorgkaneen

@ckaraneen