Learning to Select Mates in Artificial Life

Dylan Ashley*, Valliappa Chockalingam*, Braedy Kuzma*, Vadim Bulitko

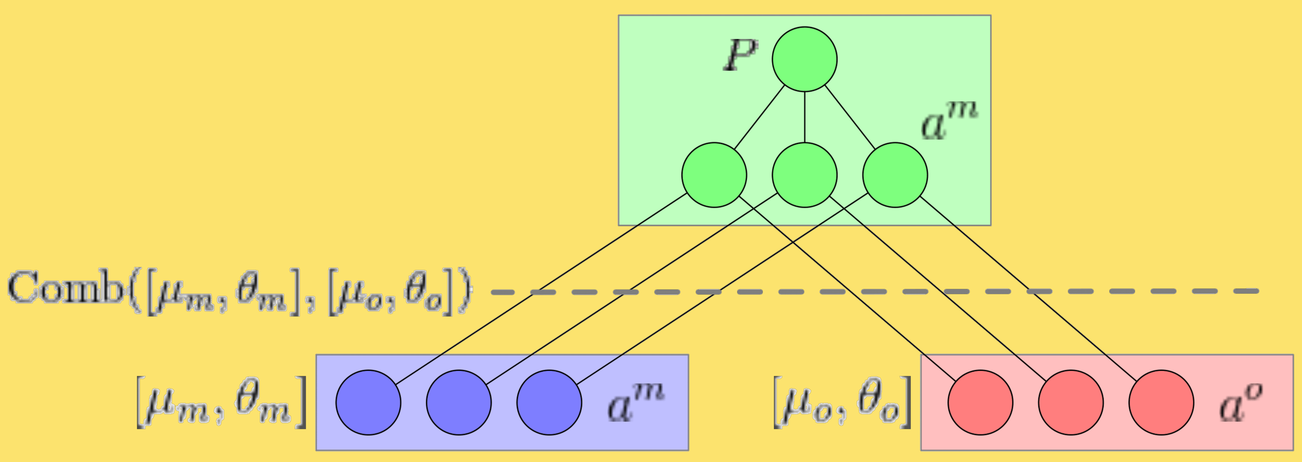
Department of Computing Science, University of Alberta, Edmononton, CA

Abstract & Motivation

- Artificial life (A-life) simulations present a natural way to study interesting phenomena emerging in a population of evolving agents.
- Allowing agents the capability to learn things on their own without human biases often improves performance (Ex: AlphaGo $\longrightarrow$ AlphaGoZero).
- Could allowing agents to select mates on their own, or in other words, making crossover learnable, extend population lifetimes?
- Each agent can evaluate potential mates with a preference function which maps information about an agent and a candidate mate to a scalar preference for deciding whether or not to form an offspring.
- Parameters of the preference function are genetically encoded in each agent, allowing preferences to be agent-specific and evolving over time.
- With a simple predator-prey A-life environment, we demonstrate that the ability to evolve a per-agent mate-selection preference function significantly increases the extinction time of the population.
- Also, an inspection of the evolved preference function parameters show, as expected, that agents evolve to favor mates who have survival traits.

Methods

- We start with an agent population A consisting of agents $\{a_1, \ldots, a_{|A|}\}$.
- Agents receive states from the state space $\mathcal{S}$ that contain agent-centric information about the environment and choose actions from the action space $\mathcal{A}$ at every timestep.
- Each agent a_i has a genome $g_i \in G$ that contains information about the agent's behaviour among other things like its appearance.
- Each genome consists of parameters for these parameterized functions:
  - An *evaluation function*, $V_\theta : \mathcal{S} \longrightarrow \mathbb{R}$,
  - An *action function*, $Q_\mu : \mathcal{S} \longrightarrow \mathbb{R}^{|\mathcal{A}|}$, and,
- As in prior Evolutionary Reinforcement Learning (ERL) work:
  - The evaluation function is the same during an agent's lifetime but is evolved over time through mutation and crossover.
  - The action function is learned through an RL algorithm and we use Q-Learning: $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a')]$ with r calculated using the evaluation function: $V(s') - V(s)$.
- Extending on the above ERL formulation, we introduce
  - A *combination function* Comb $: G \times G \longrightarrow C$ that maps pairs of genomes to information about how the genomes relate.
  - A parameterized *preference function* to each genome, $\rho_\nu : C \longrightarrow [0, 1]$ that uses a sigmoid squashing function for the rescaling.

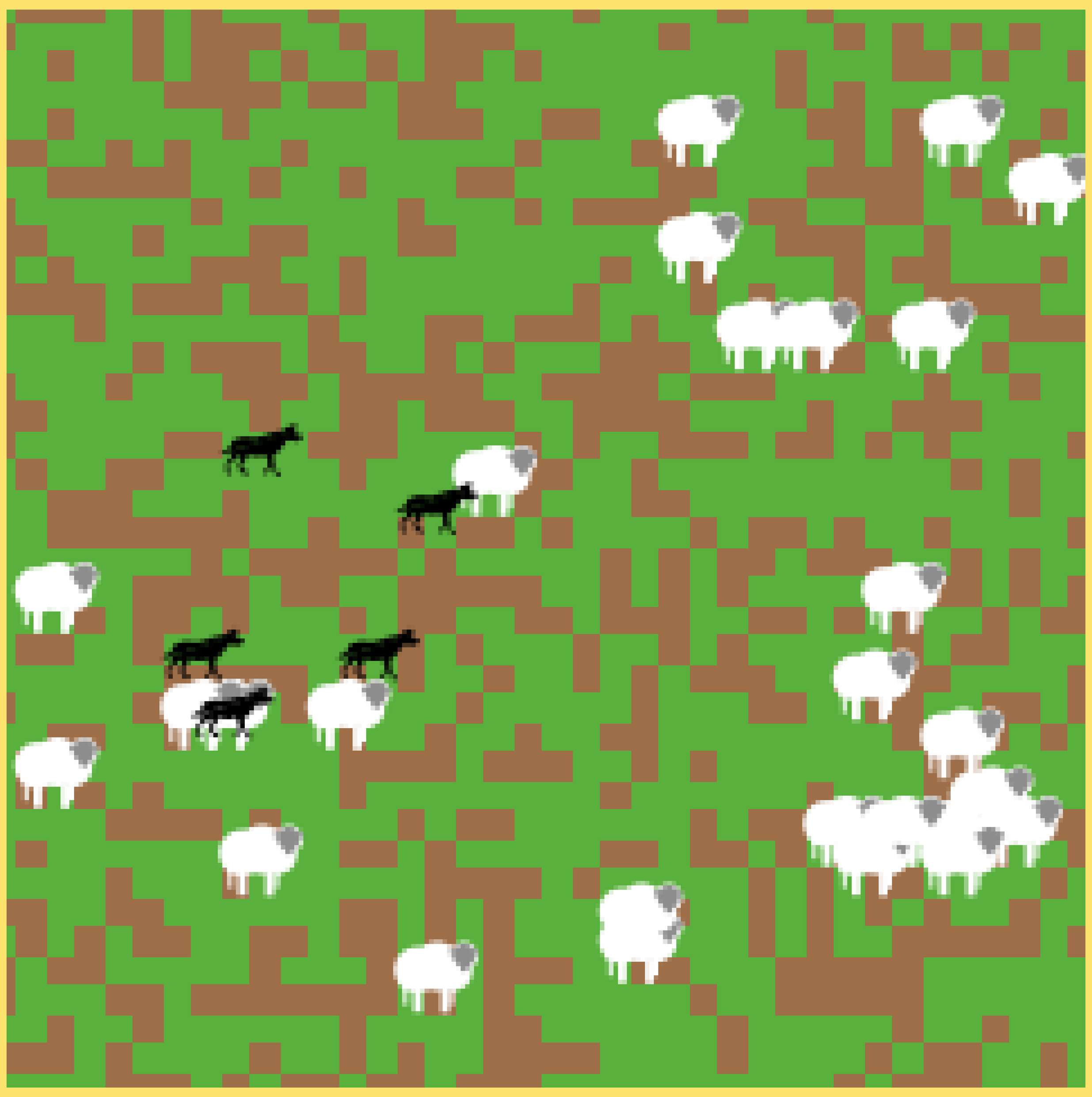


- The preference function is also kept the same during an agent's lifetime but evolved over time using mutation and crossover.
- We test multiple combination functions: Euclidean distance, element-wise squared distance, element-wise absolute difference, and the identity transformation which simply returns the other agent's genome.

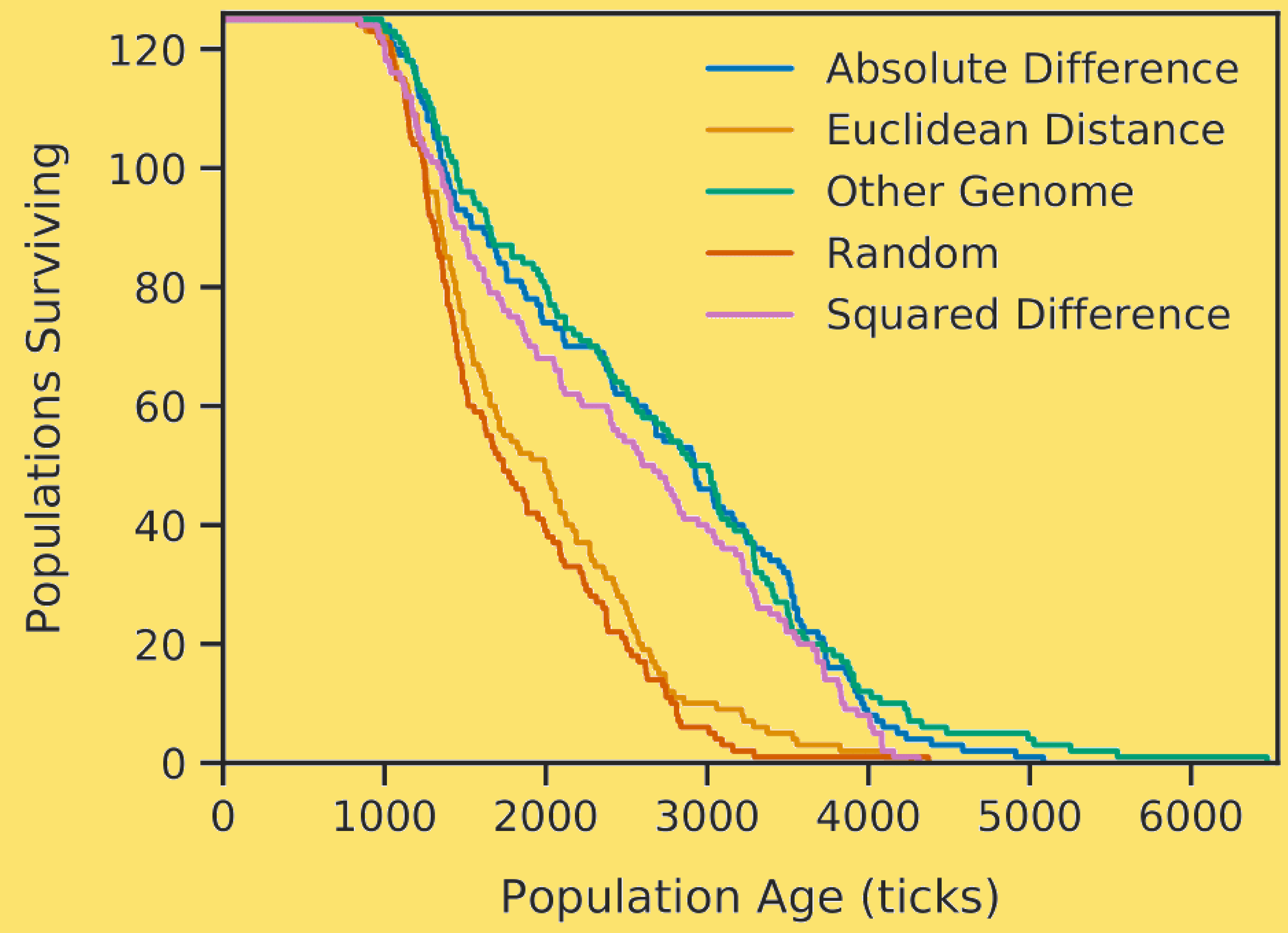
Experiments

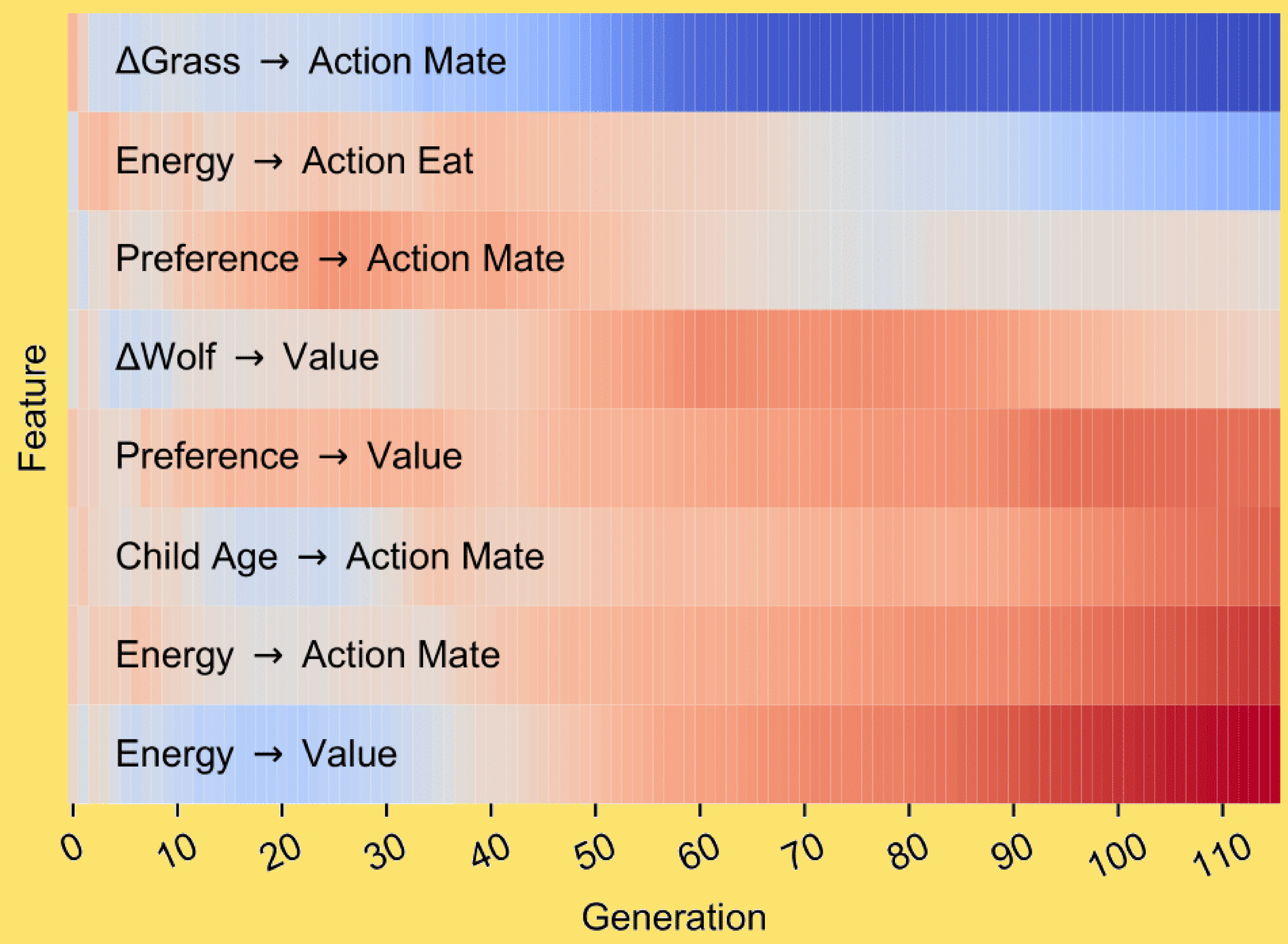
- We use a wolf-sheep predation model in NetLogo where the sheep need to eat grass to survive, the wolves need to eat sheep to survive, and grass regenerates over time.
- The sheep learn through the methods proposed above and the wolves follow a fixed policy of finding and pursuing close sheep.
- The actions available to the sheep are: moving in the four cardinal directions, mating and eating grass.
- The states given to the sheep are nine-tuples: $(E, \theta_{\text{Sheep}}, \mathbf{\Delta}_{\text{Sheep}}, \theta_{\text{Wolf}}, \mathbf{\Delta}_{\text{Wolf}}, \theta_{\text{Grass}}, \mathbf{\Delta}_{\text{Grass}}, P, A)$ corresponding to Energy, Distances and Angles to closest sheep, wolf and grass, Preference scalar using the closest sheep, and the oldest age of an agent's child that the agent has encountered, a useful proxy for how well the agent is doing.

Domain



Results





Discussion

- Smarter crossover automatically allows agents to choose mates based on desired diversification or intensification of traits.
- Connection to Meta-Learning: Learning a preference function encourages agents to learn to prefer agents who have good discerning capabilities with the common goal of constructing a population full of high fitness agents, thus extending population lifetimes.

Note

An Extended Version of this paper is to be presented at IEEE Conference on Games, 2019 in London, UK.

Contact: valliapp@ualberta.ca