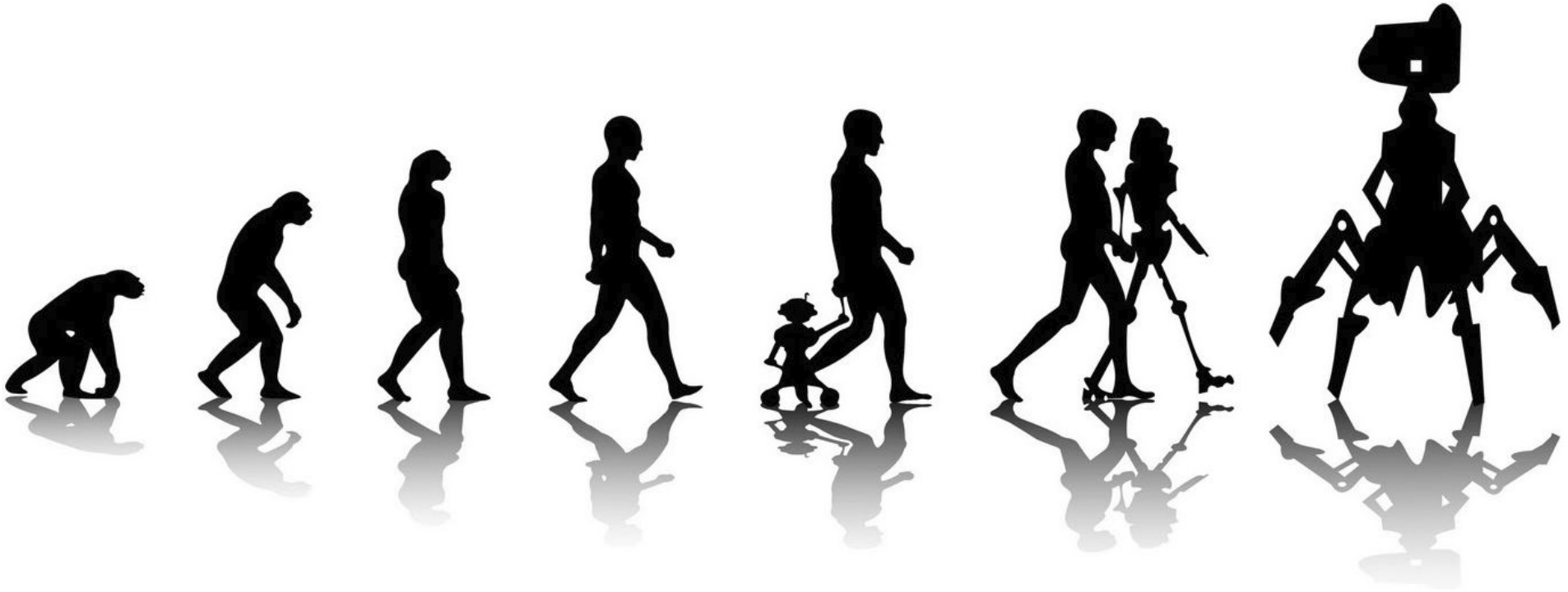


Modern Robotics: Evolutionary Robotics

COSC 4560 / COSC 5560

Professor Cheney

3/9/18

Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images

Anh Nguyen
University of Wyoming
anguyen8@uwyo.edu

Jason Yosinski
Cornell University
yosinski@cs.cornell.edu

Jeff Clune
University of Wyoming
jeffclune@uwyo.edu

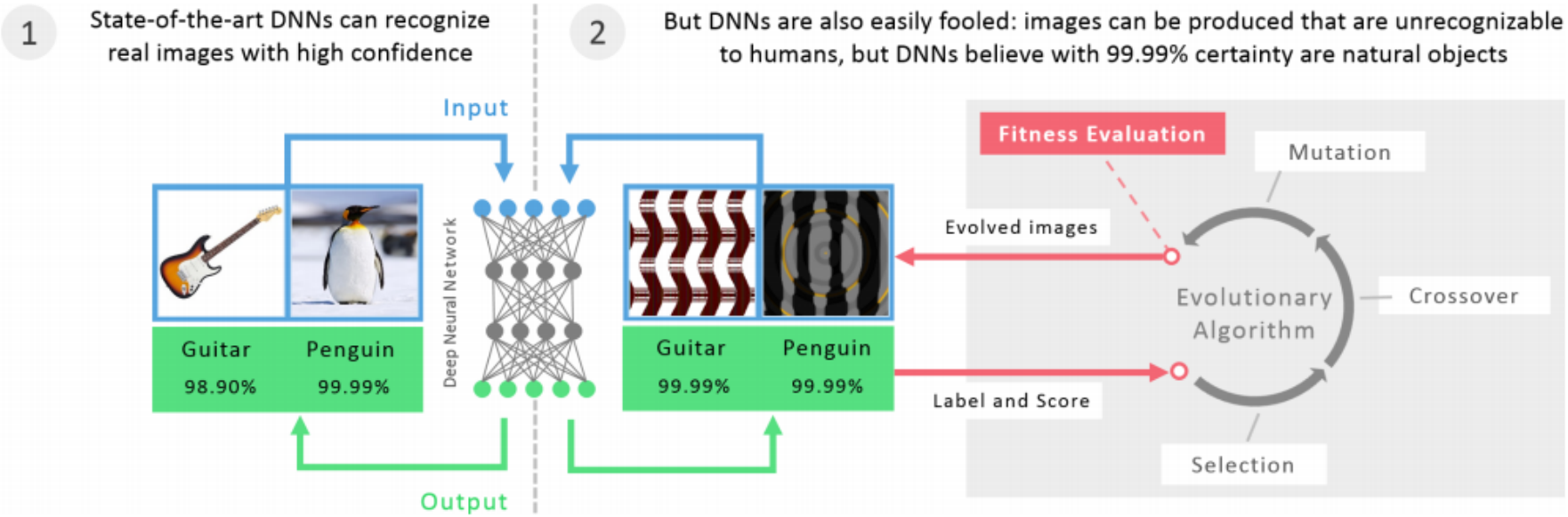


Figure 2. Although state-of-the-art deep neural networks can increasingly recognize natural images (*left panel*), they also are easily fooled into declaring with near-certainty that unrecognizable images are familiar objects (*center*). Images that fool DNNs are produced by evolutionary algorithms (*right panel*) that optimize images to generate high-confidence DNN predictions for each class in the dataset the DNN is trained on (here, ImageNet).

Direct encoding:

CPPN encoding:

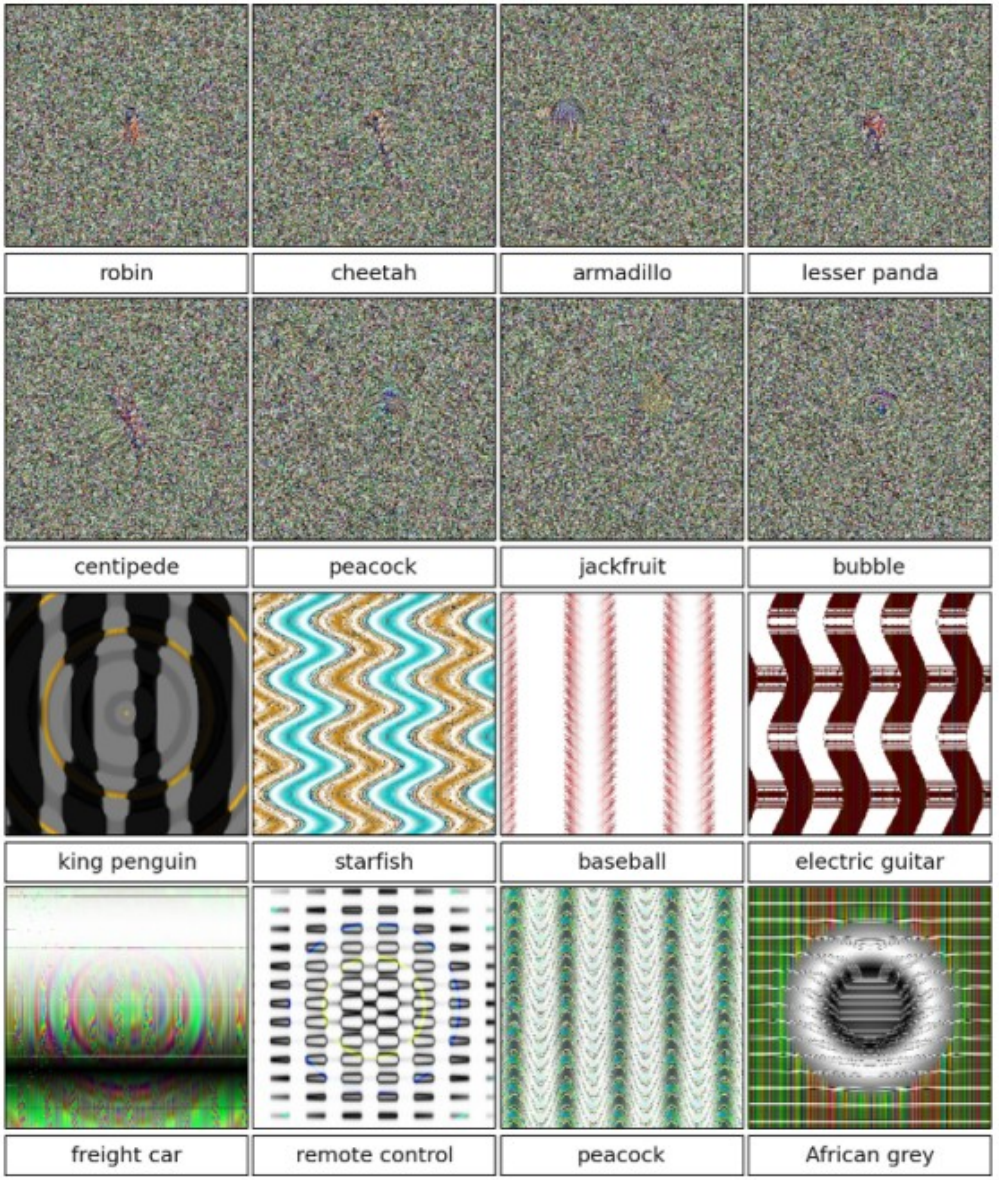


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

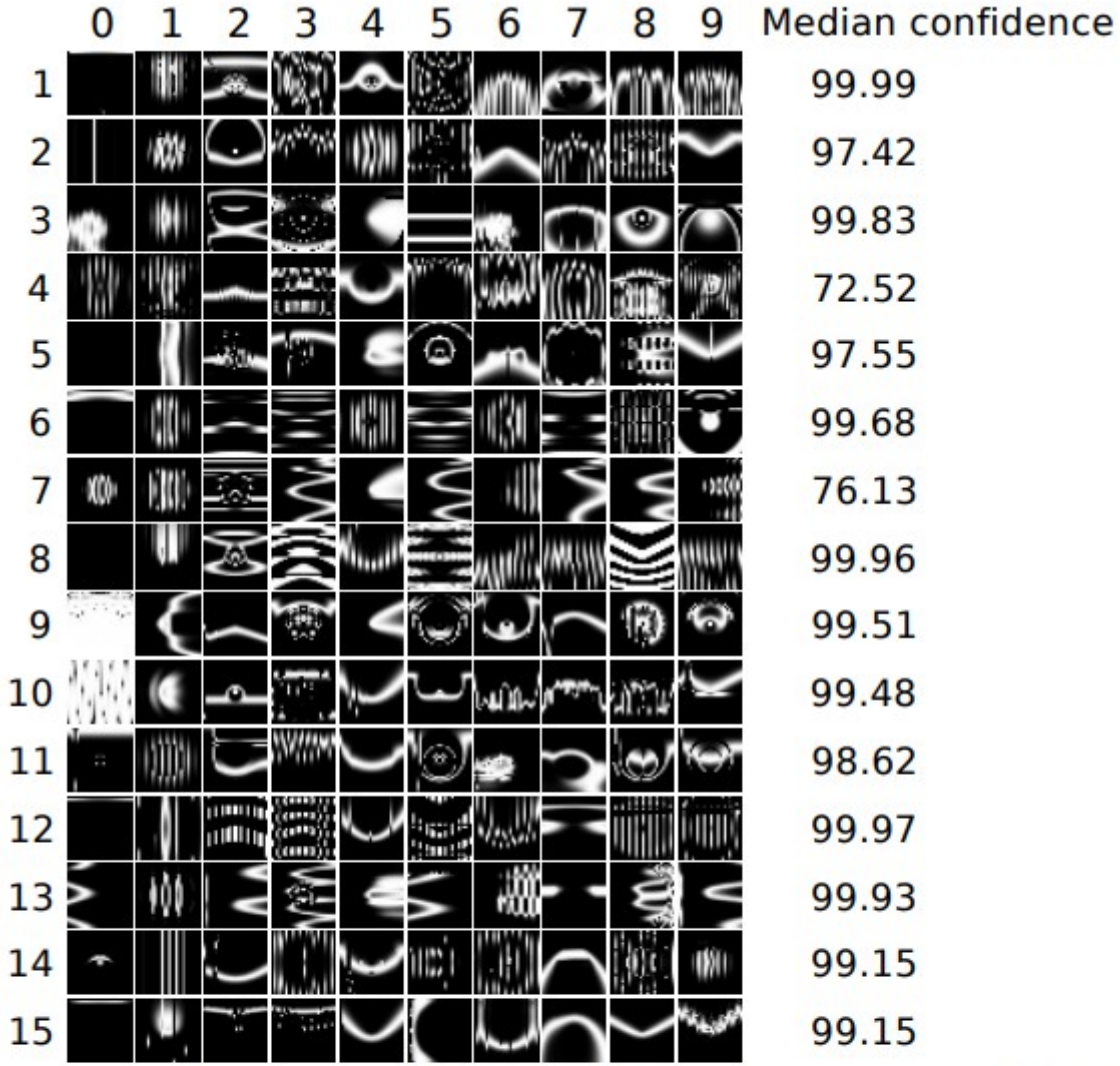


Figure 11. Training MNIST DNN_i with images that fooled MNIST DNN_1 through DNN_{i-1} does not prevent evolution from finding new fooling images for DNN_i. Columns are digits. Rows are DNN_i for $i = 1...15$. Each row shows the 10 final, evolved images from one randomly selected run (of 30) per iteration. Medians are taken from images from all 30 runs.

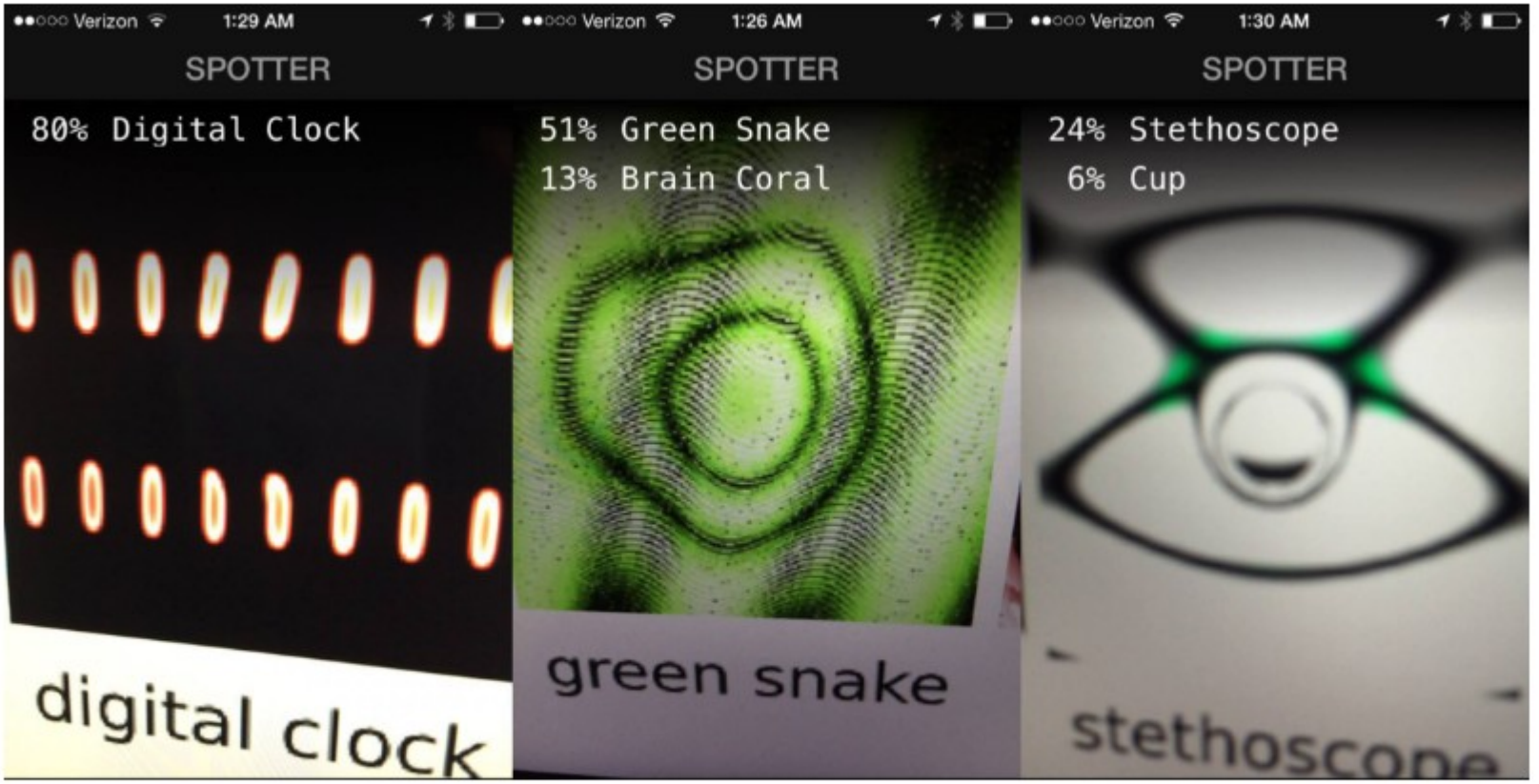


Figure 6: Dileep George told us (via Alexander Terekhov) that he pointed an image recognition iPhone app powered by Deep Learning at our "fooling images" displayed on a computer screen and the iPhone/app was equally fooled! That's very interesting given the different lighting, angle, camera lens, etc. It shows how robustly the DNN feels these images are the genuine articles.

Multi-Agent Systems

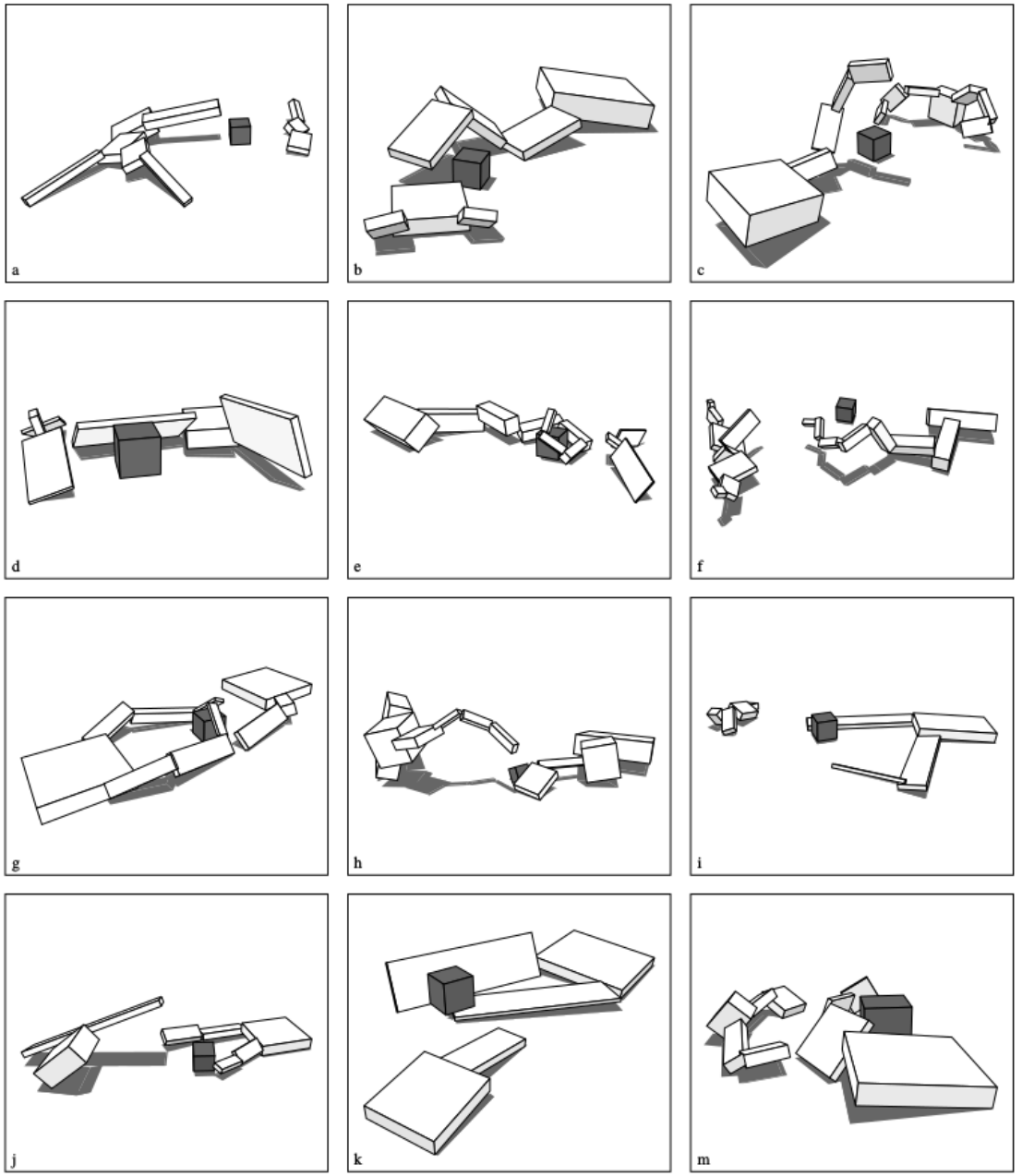


Figure 9: Evolved competing creatures.

Approximating Competitive Environments

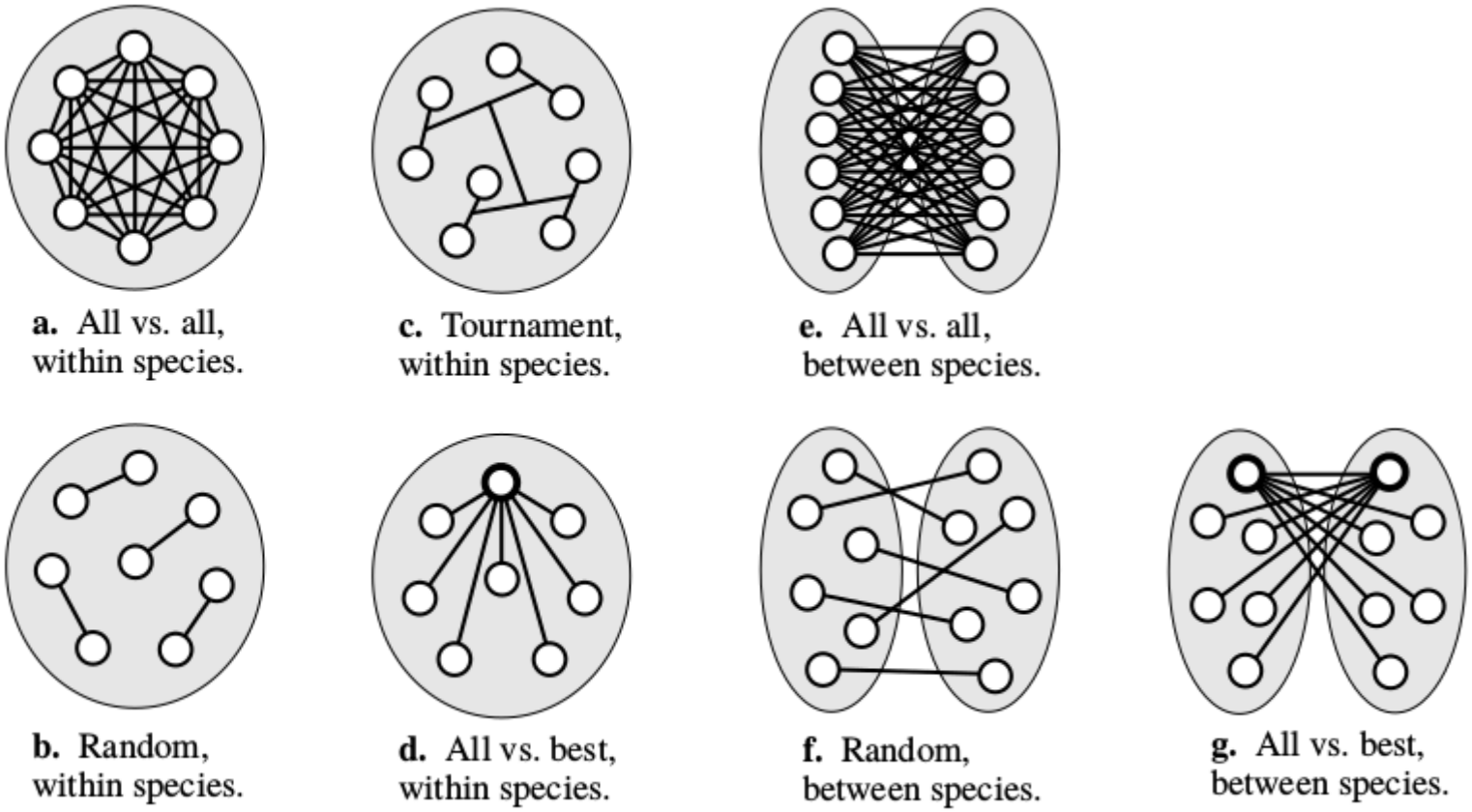


Figure 2: Different pair-wise competition patterns for one and two species. The gray areas represent species of interbreeding individuals, and lines indicate competitions performed between individuals.

Evolution of Swarming Behavior Is Shaped by How Predators Attack

Randal S. Olson , David B. Knoester and Christoph Adami

Posted Online August 17, 2016
https://doi.org/10.1162/ARTL_a_00206


natgeotv.com

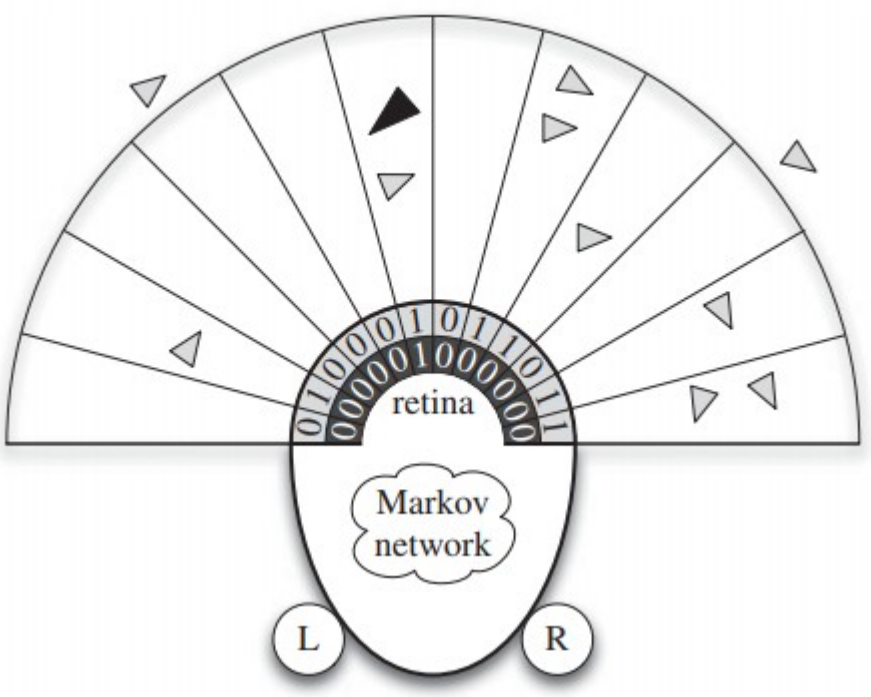
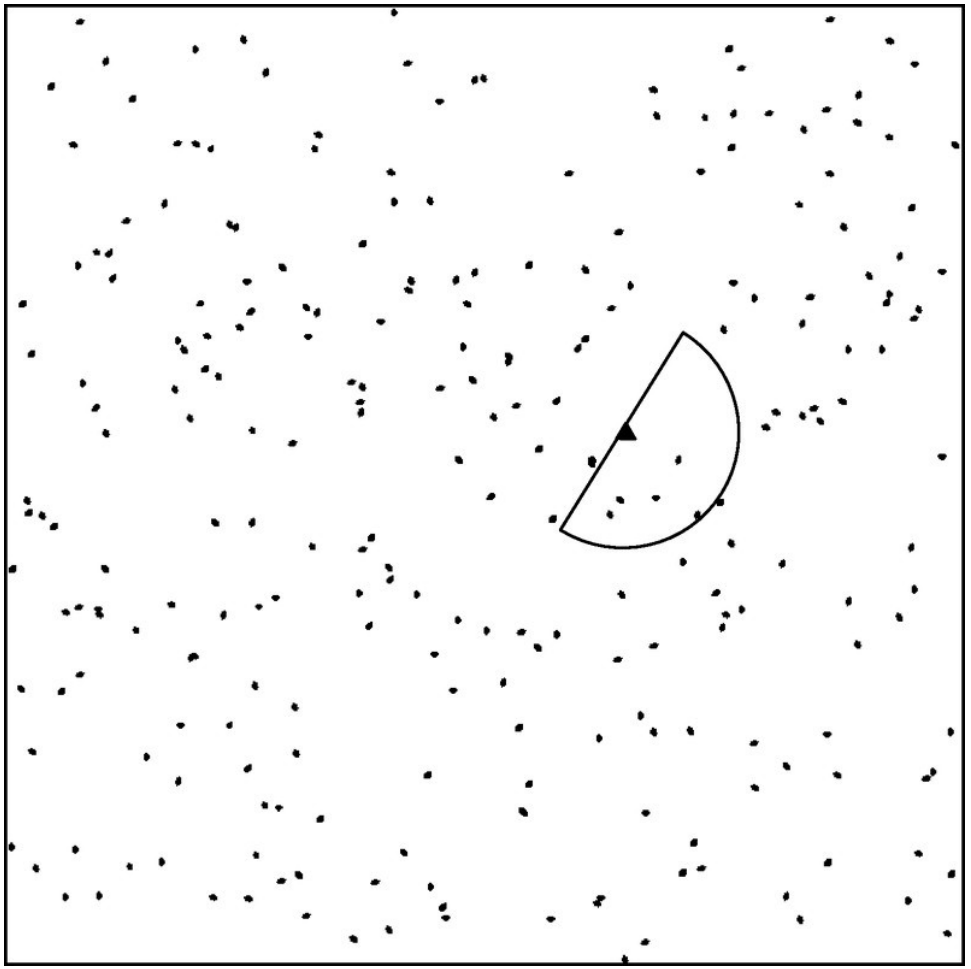
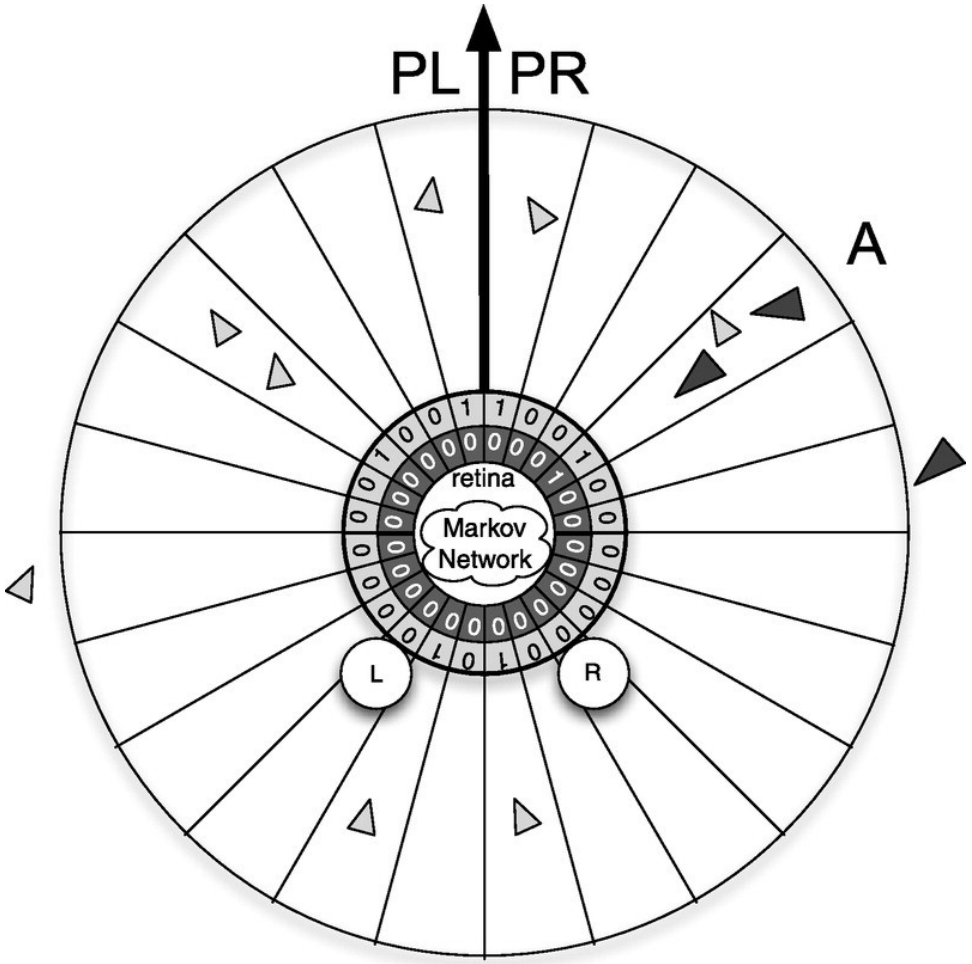


Figure 1. An illustration of the predator and prey agents in the model. Light grey triangles are prey agents and the dark grey triangle is a predator agent. The predator and prey agents have a 180° limited-distance retina (100 virtual metres for the prey agents; 200 virtual metres for the predator agent) to observe their surroundings and detect the presence of the predator and prey agents. Each agent has its own Markov network, which decides where to move next based on a combination of sensory input and memory. The left and right actuators (labelled 'L' and 'R') enable the agents to move forward, left, and right in discrete steps.




PL
PR
A
retina
Markov
Network
L
R

(a)

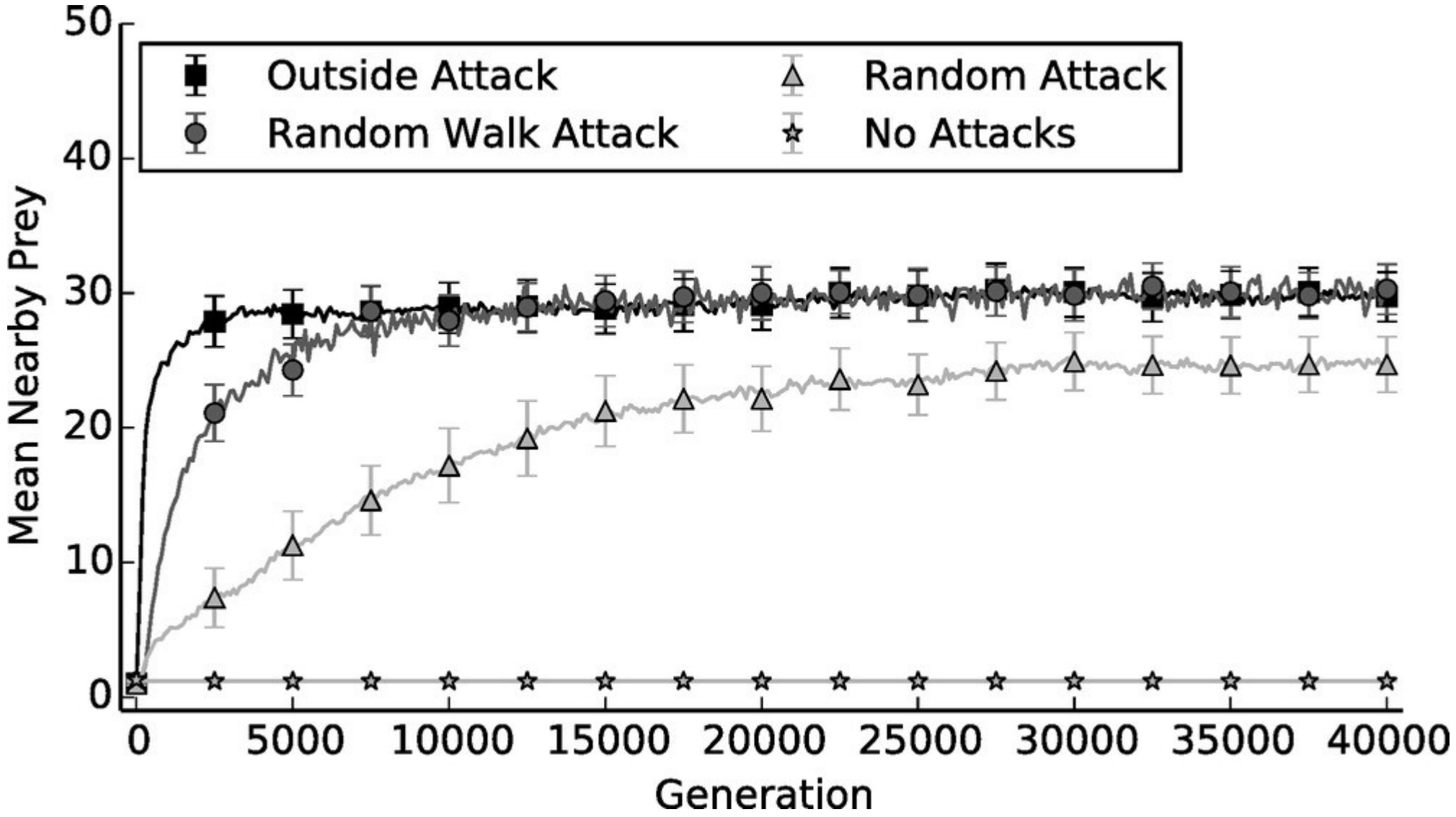
Random Attacks

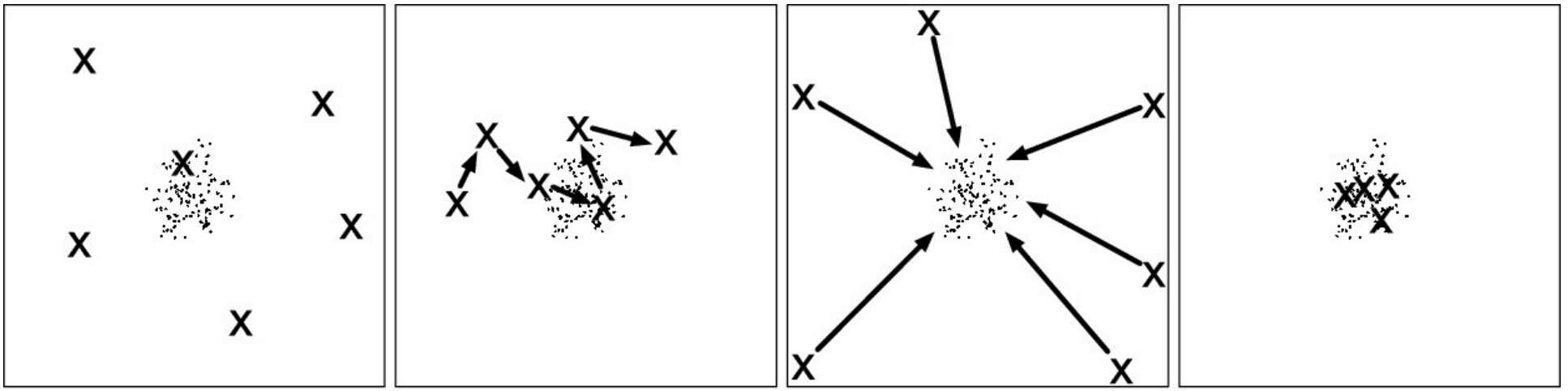
(b)

Random Walk Attacks

(c)

Outside Attacks


Outside Attack
Random Attack
Random Walk Attack
No Attacks
Mean Nearby Prey
Generation
0
10
20
30
40
50
0
5000
10000
15000
20000
25000
30000
35000
40000

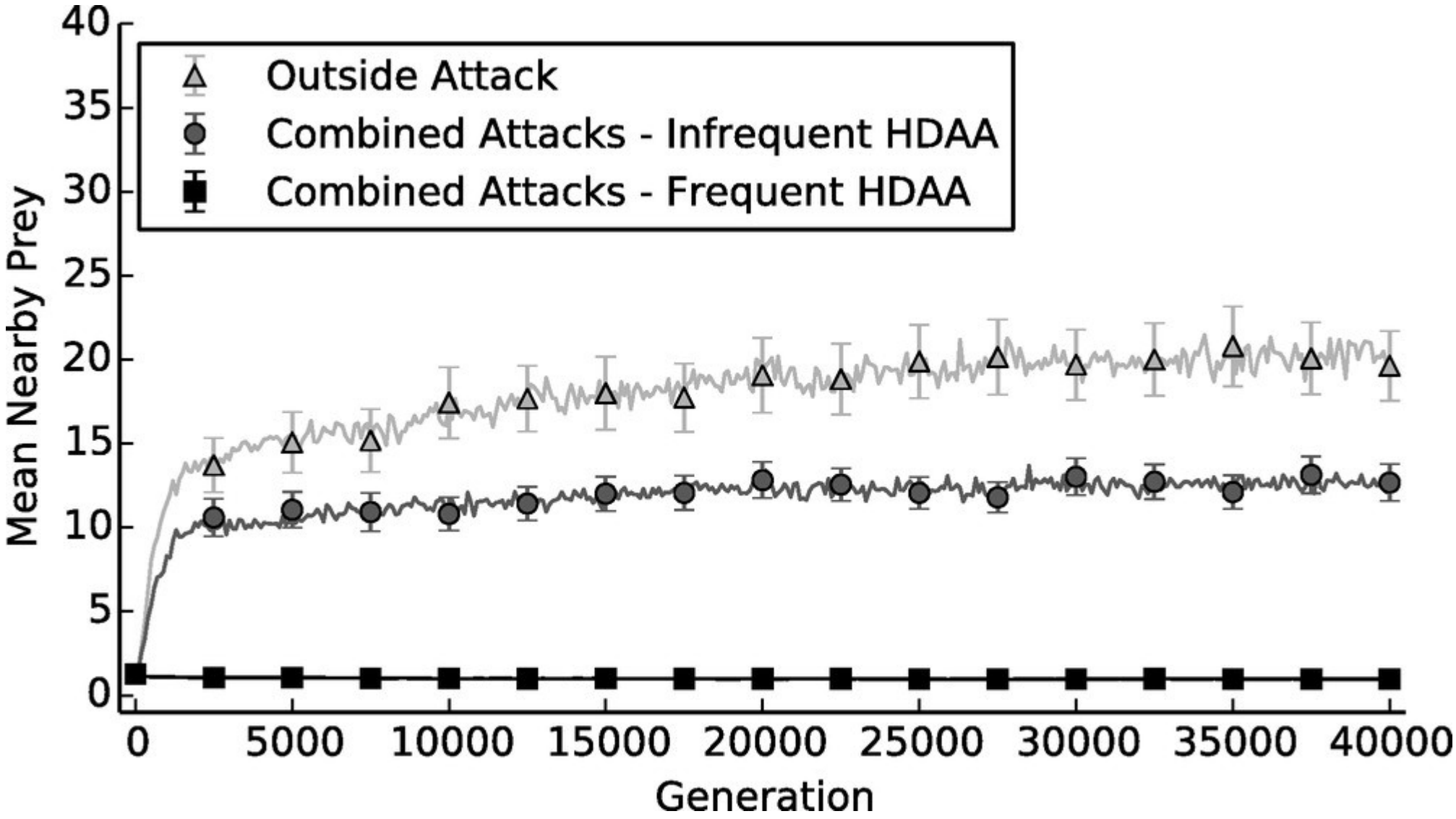


(a) Random Attacks

(b) Random Walk Attacks

(c) Outside Attacks

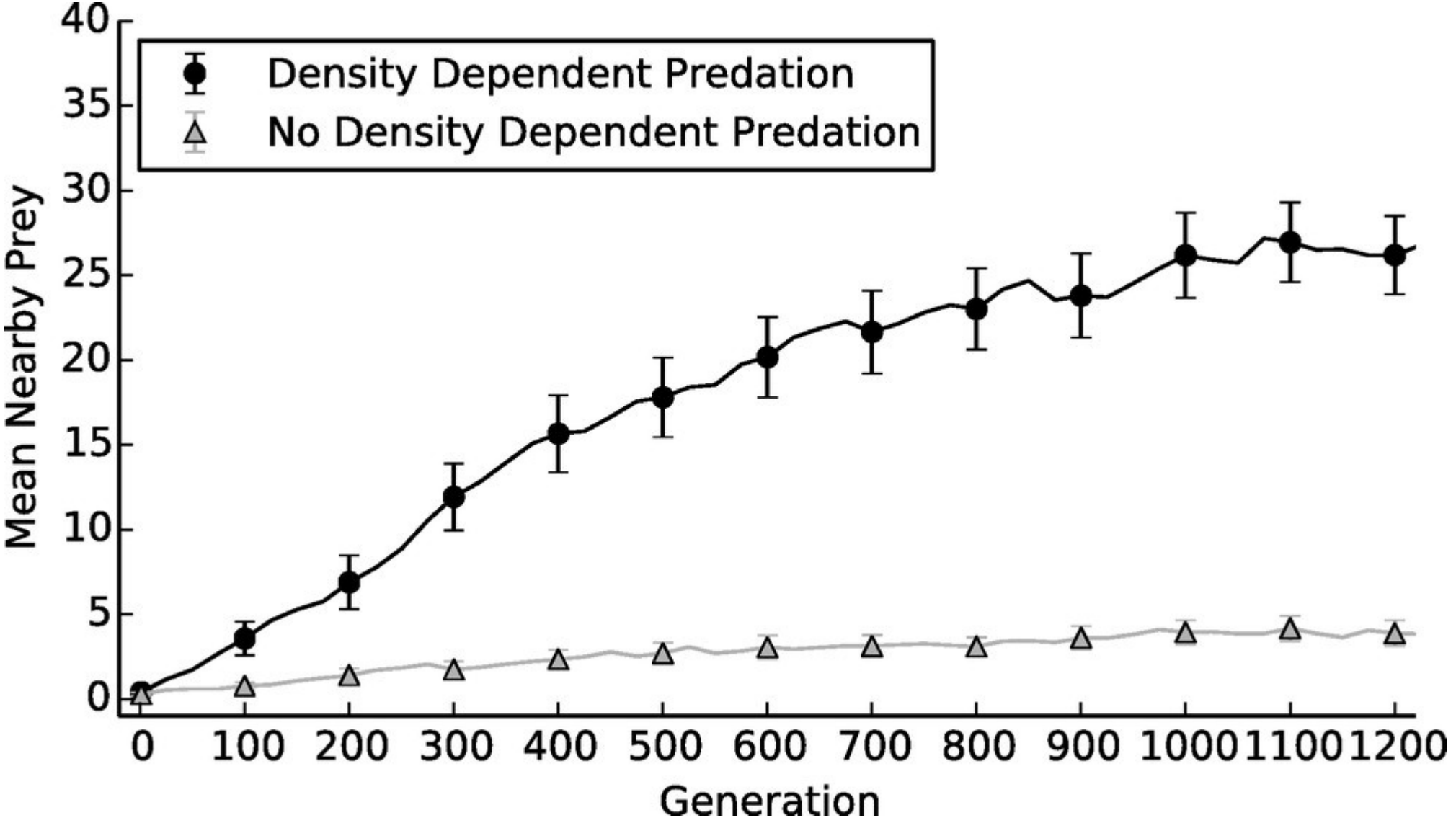
(d) High-Density Area Attacks

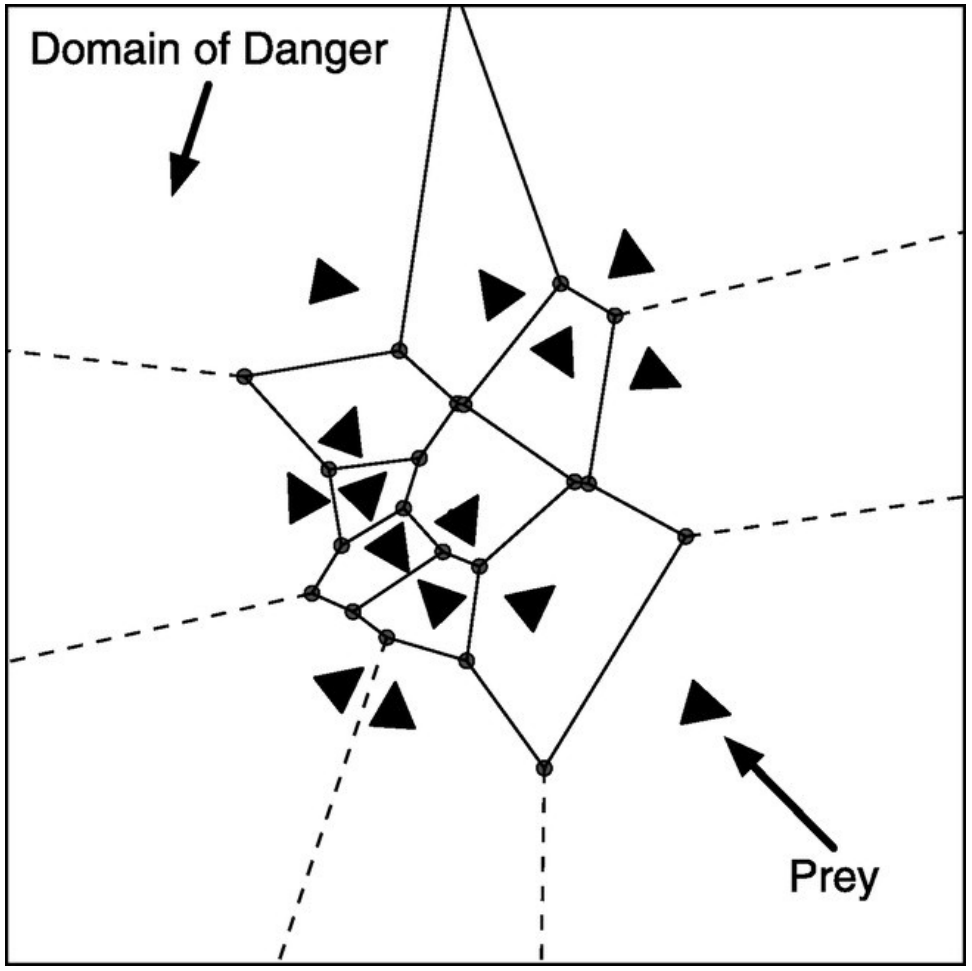

Outside Attack
Combined Attacks - Infrequent HDAA
Combined Attacks - Frequent HDAA
Mean Nearby Prey
Generation
0
5
10
15
20
25
30
35
40
0
5000
10000
15000
20000
25000
30000
35000
40000

Density-Dependent Predation

$$P_{\text{capture}} = \frac{1}{A_{\text{NV}}},$$

where A_{NV} is the number of prey agents that are visible to the predator, i.e. anywhere in the predator agent's visual field, and within 30 virtual metres of the target prey.


Density Dependent Predation
No Density Dependent Predation
Mean Nearby Prey
Generation
0
5
10
15
20
25
30
35
40
0
100
200
300
400
500
600
700
800
900
1000
1100
1200


Domain of Danger
Prey

Here we see that density-dependent predation provides a sufficient selective advantage for prey to evolve the selfish herd in response to predation by coevolving predators, despite the fact that swarming prey experience an increased attack rate from the predators due to this behavior

Accordingly, these results uphold Hamilton's hypothesis that grouping behavior could evolve in animals purely for selfish reasons, without the need for an explanation that involves the benefits to the whole group [18]. Moreover, the discoveries in this work refine the selfish herd hypothesis by clarifying the effect that different attack modes have on the evolution of the selfish herd.

Predator confusion is sufficient to evolve swarming behaviour

Randal S. Olson[1,4], Arend Hintze[2,4], Fred C. Dyer[3,4], David B. Knoester[2,4] and Christoph Adami[2,4]

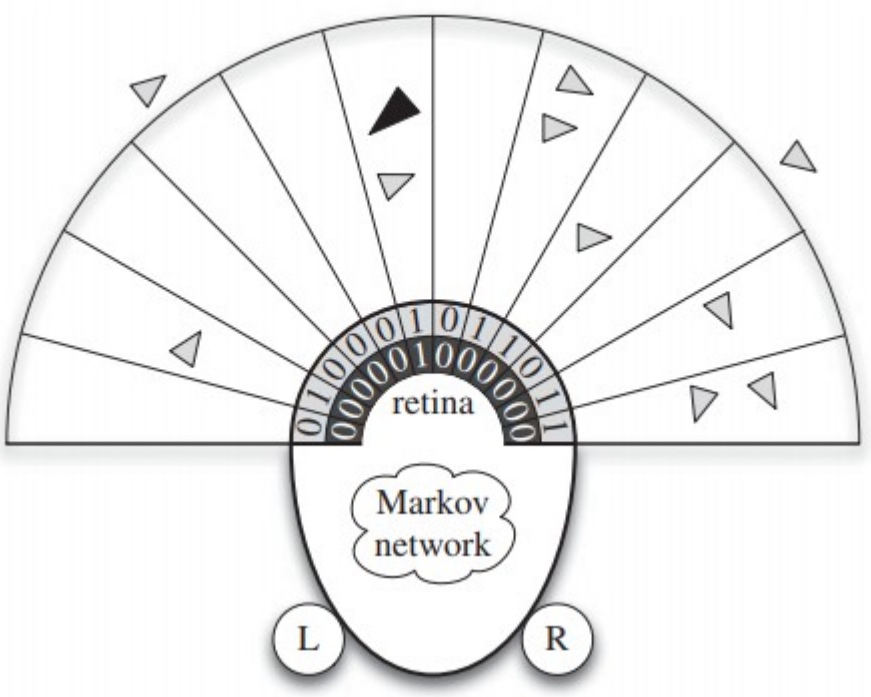


Figure 1. An illustration of the predator and prey agents in the model. Light grey triangles are prey agents and the dark grey triangle is a predator agent. The predator and prey agents have a 180° limited-distance retina (100 virtual metres for the prey agents; 200 virtual metres for the predator agent) to observe their surroundings and detect the presence of the predator and prey agents. Each agent has its own Markov network, which decides where to move next based on a combination of sensory input and memory. The left and right actuators (labelled 'L' and 'R') enable the agents to move forward, left, and right in discrete steps.

we assign the predator and prey genomes separate fitness values according to the fitness functions:

$$W_{\text{predator}} = \sum_{t=1}^{2000} S - A_t$$

and

$$W_{\text{prey}} = \sum_{t=1}^{2000} A_t,$$

where t is the current simulation time step, S is the starting swarm size (here, $S = 50$), and A_t is the number of prey agents alive at simulation time step t.

In our coevolution experiments, the predator agents can detect only nearby prey agents, using a limited-distance (200 virtual meters), pixelated retina covering its frontal 180° that works just like the prey agent's retina (Figure 3). Similarly to the prey agents, predators make decisions about how to move next using their MN, as shown in Table 1, but move 3 times faster than the prey agents and turn correspondingly slower (6° per simulation time step) due to their higher speed. This dramatically faster predator movement speed is meant to represent predators that perform rapid attacks on groups of prey, such as a peregrine falcon dive-bombing a swarm of starlings. Finally, if a predator agent moves within 5 virtual meters of a prey agent that is anywhere within its retina, the predator agent attempts an attack on the prey agent. If the attempt is successful, we remove the prey agent from the simulation and mark it as consumed.

Density-Dependent Predation

$$P_{\text{capture}} = \frac{1}{A_{\text{NV}}},$$

where A_{NV} is the number of prey agents that are visible to the predator, i.e. anywhere in the predator agent's visual field, and within 30 virtual metres of the target prey.

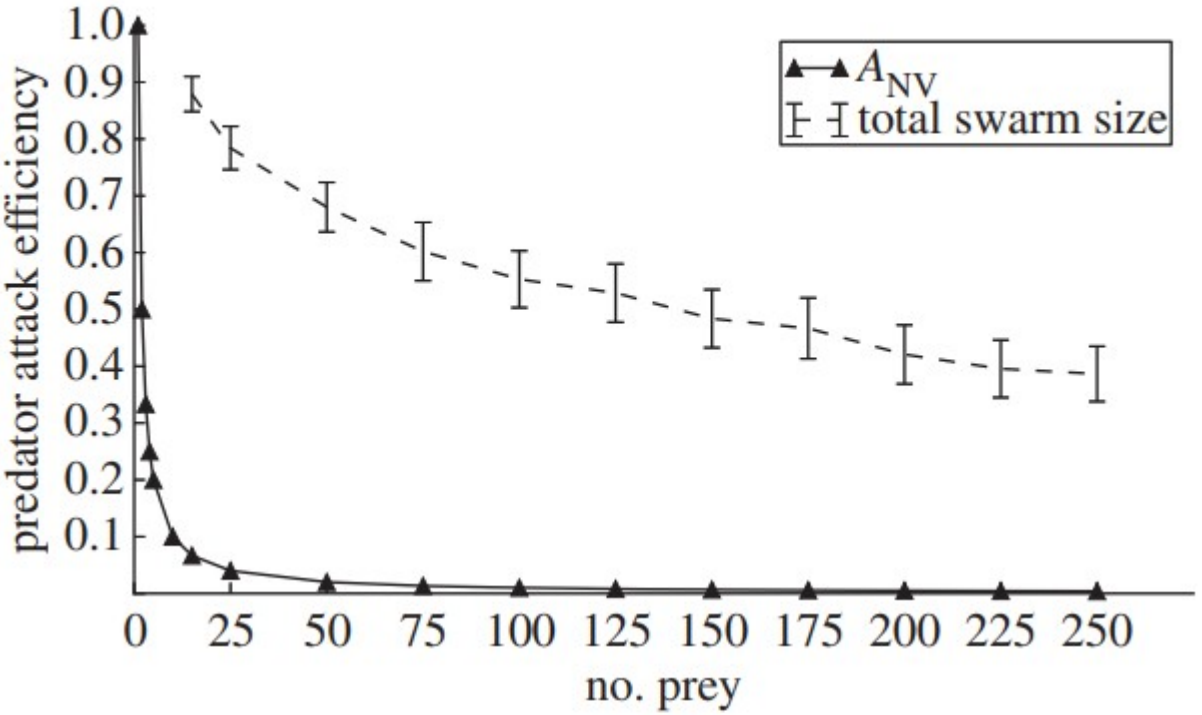


Figure 2. Relation of predator attack efficiency (no. of successful attacks/total no. of attacks) to number of prey. The solid line with triangles indicates predator attack efficiency as a function of the number of prey within the visual field of the predator (A_{NV}). Similarly, the dashed line with error bars shows the actual predator attack efficiency given the predator attacks a group of swarming prey of a given size, using the A_{NV} curve to determine the per-attack predator attack success rate. Error bars indicate two standard errors over 100 replicate experiments.

dispersion

reformation

shape change

circular groups

chaotic swarm

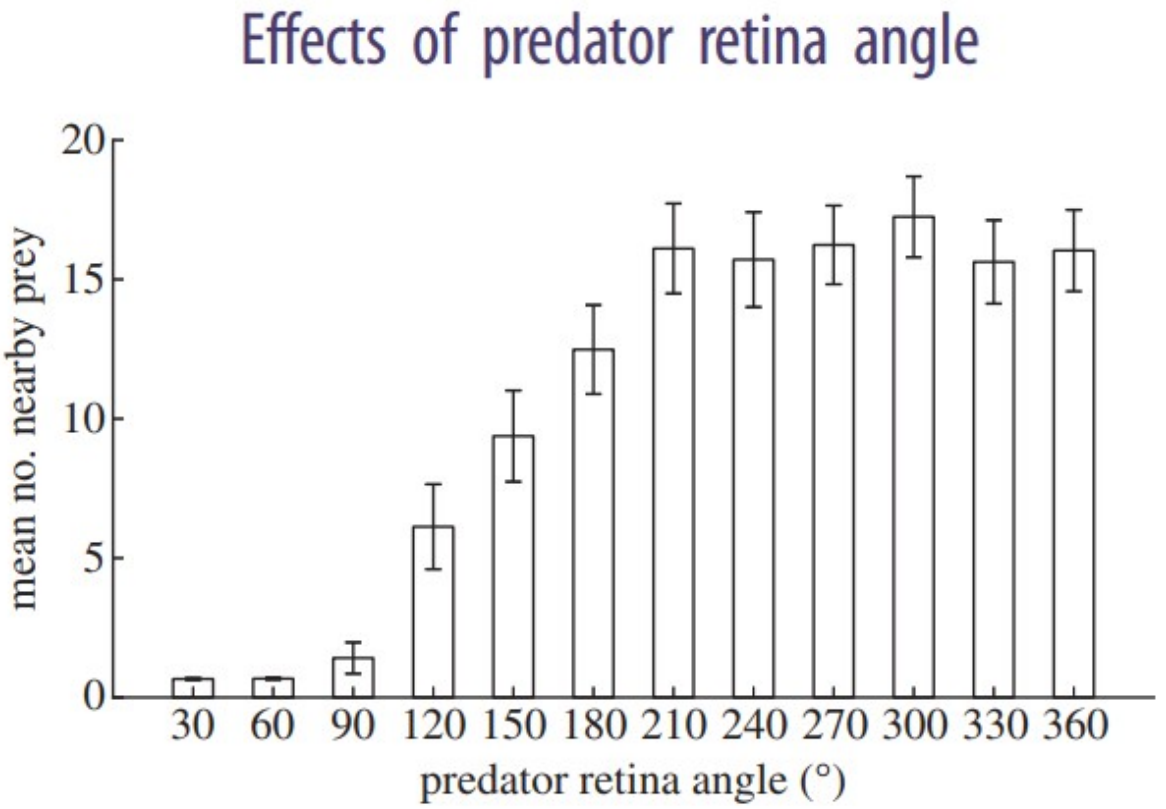


Figure 5. Mean swarm density at generation 1200 as a function of predator view angle. Swarming to confuse the predator was an ineffective behaviour if the predator's visual field covered only the frontal 60° or less, owing to the predator's focused retina. As the predator's visual field was incrementally increased to cover the frontal 90° and beyond, predator confusion via swarming again became an effective anti-predator behaviour, as evidenced by the swarms exhibiting significantly higher swarm density at generation 1200. Error bars indicate 2 s.e. across 180 replicate experiments.

Coevolution of Role-Based Cooperation in Multi-Agent Systems

Chern Han Yong and Risto Miikkulainen

evaluated in a task where a team of several predators must cooperate to capture a fast-moving prey.

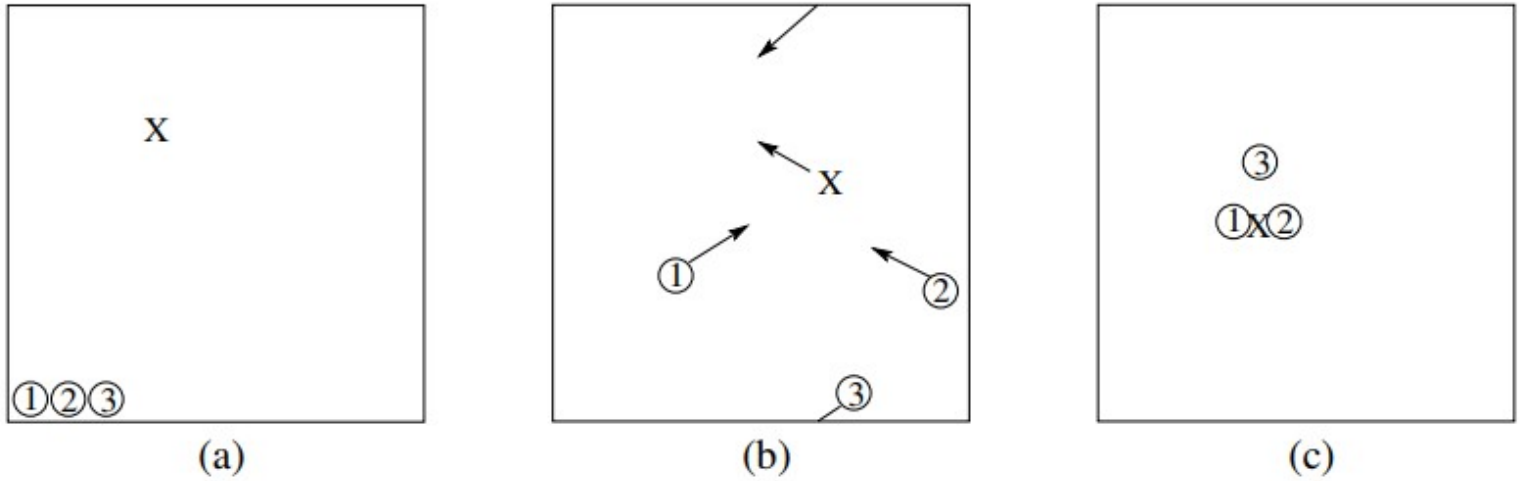


Fig. 2. **The Prey-Capture Task.** The environment is a 100×100 toroidal grid, with one prey (denoted by "X") and three predators (denoted by "1", "2" and "3"). Figure (a) illustrates a starting scenario: The predators start in a row at the bottom left corner, and the prey starts in a random location. Figure (b) illustrates a scene later during a trial. The arrows indicate a general direction of movement: Since each agent may only move in the four cardinal directions, a movement arrow pointing 45 degrees northwest means the agent is moving north and west on alternate time steps. Figure (c) shows the positions of the predators one time step before a successful capture. The prey always moves directly away from the nearest predator; even though it is as fast as the predators, if the predators approach it consistently from different directions, eventually the prey has nowhere to run.

Cooperative Coevolution With vs. Without Communication

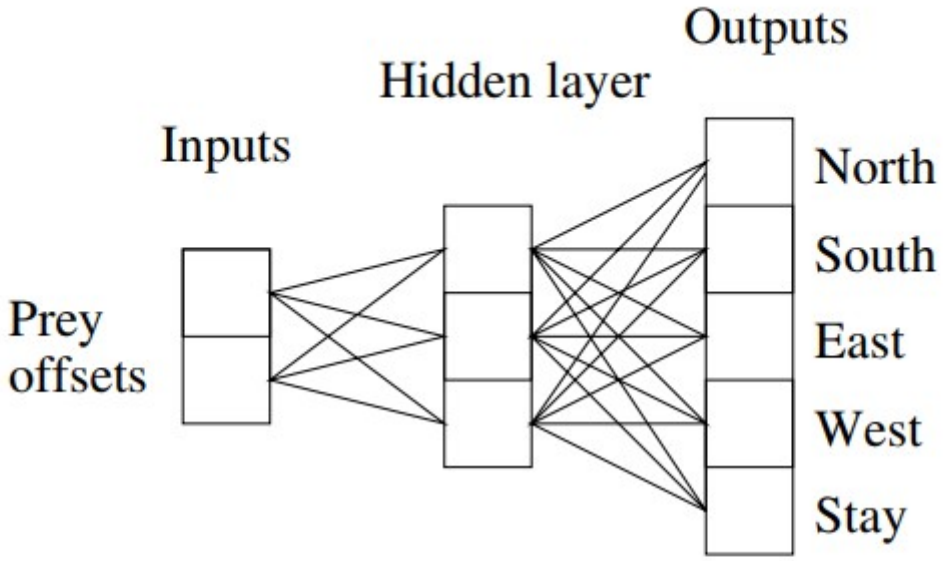


Fig. 6. **Controller for each autonomous non-communicating predator.** This network receives the prey's x and y offsets as its inputs. Therefore, it controls a single predator without knowing where the other two predators are (i.e. there is no communication between them). There are three hidden units, and the chromosomes for each hidden layer unit consist of seven real-valued numbers (two inputs and five outputs).

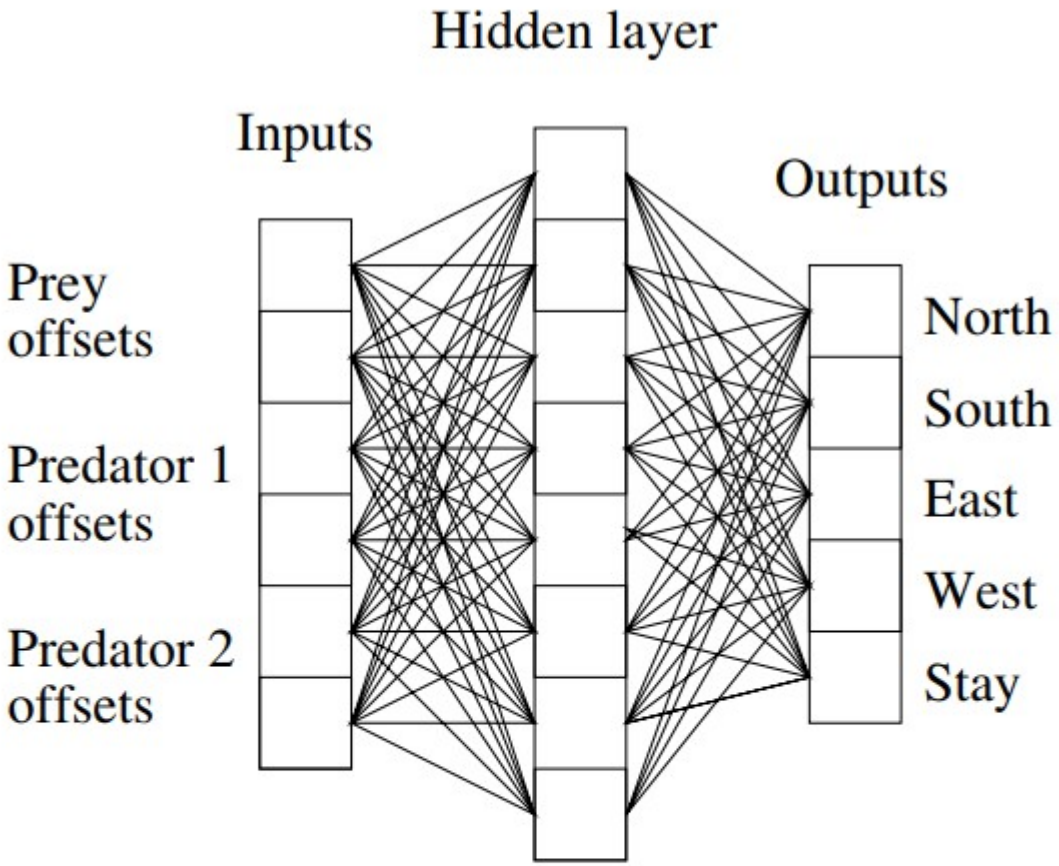


Fig. 5. **Controller for each autonomous communicating predator.** This network autonomously controls one of the predators; three such networks are simultaneously evolved in the task. The locations of this predator's teammates are obtained, and their relative x and y offsets are calculated and given to this network as information obtained through communication. It also receives the x and y offsets of the prey. There are eight hidden units, and the chromosomes for each hidden layer unit consist of 11 real-valued numbers (six inputs and five outputs).

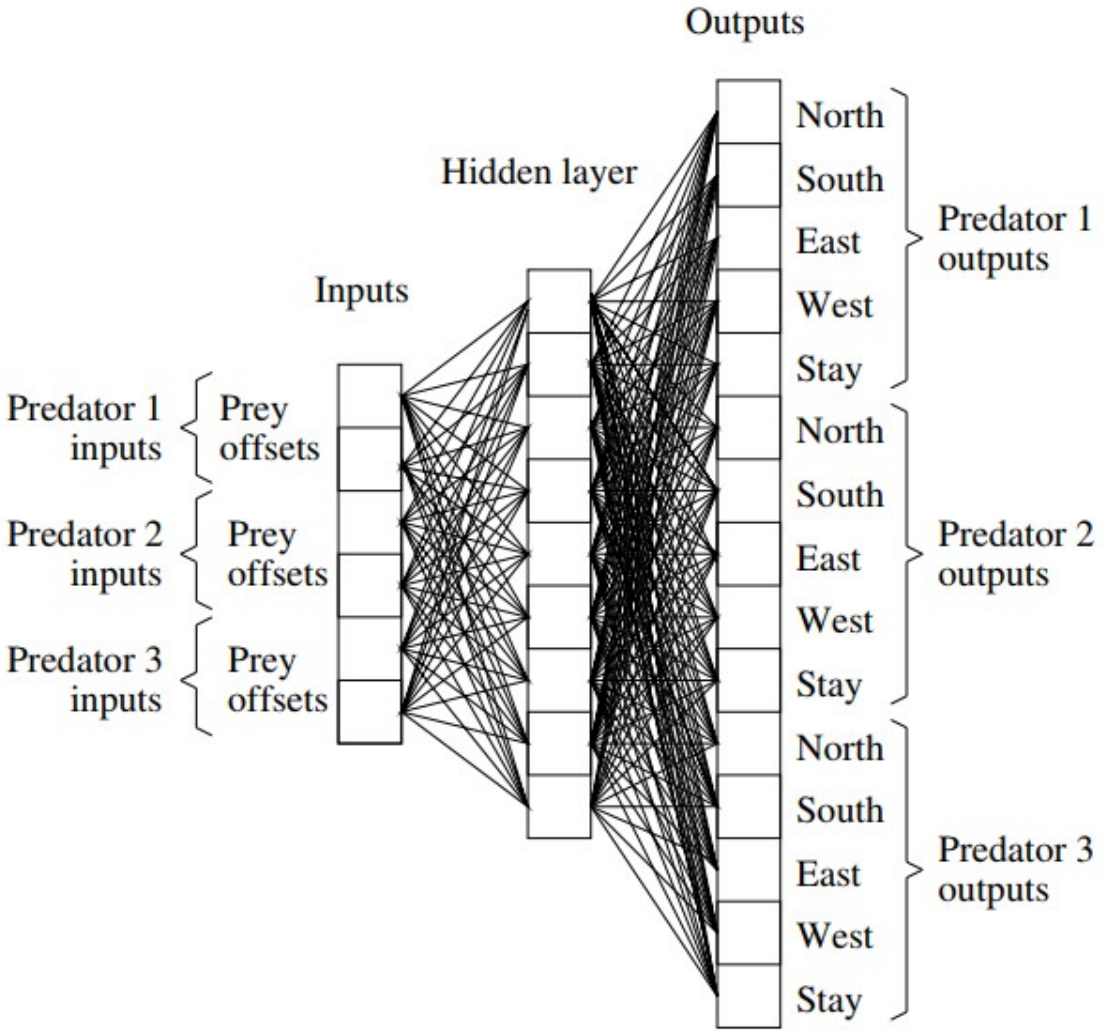


Fig. 4. **Central controller network for a team of three predators.** This network receives the relative x and y offsets (i.e. relative distance) of the prey from the perspective (i.e. location) of all three predators, and outputs the movement decisions for all three predators. This way it acts as the central controller for the whole team. There are nine hidden units, and the chromosomes for each hidden layer unit consist of 21 real-valued numbers (six inputs and 15 outputs).

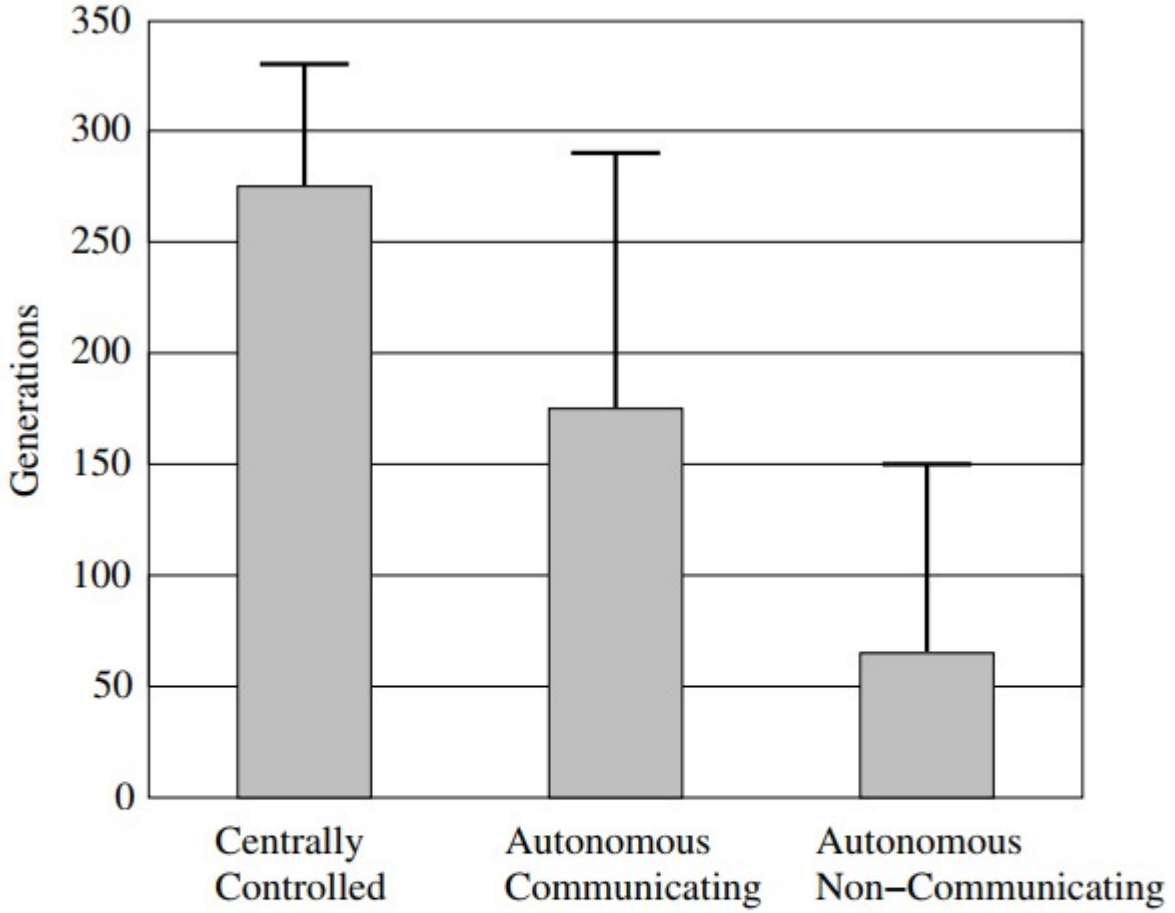


Fig. 7. **Evolution performance for each approach.** The average number of generations, with standard deviation, required to solve the task is shown for each approach. The centrally controlled team took 50% longer than the autonomously controlled communicating team, which in turn took over twice as long as the autonomously controlled non-communicating team, to evolve a successful solution. All differences are statistically significant ($p < 0.05$).

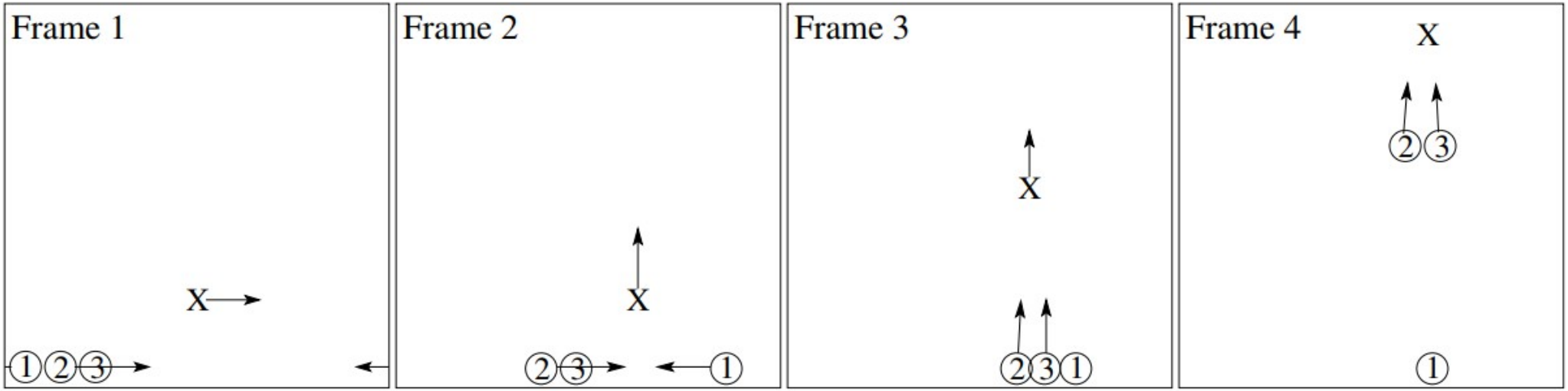


Fig. 10. **A sample strategy of a non-communicating team.** In frames 1 and 2, the predators are in setup mode, maneuvering into an appropriate chase configuration. In frame 3, they switch to chase mode: Predators 2 and 3 chase the prey toward predator 1, which acts as a blocker. This strategy is effective and does not require communication. Animated demos of this strategy, and others discussed in this paper, are available at http://nn.cs.utexas.edu/?multiagent-esp.

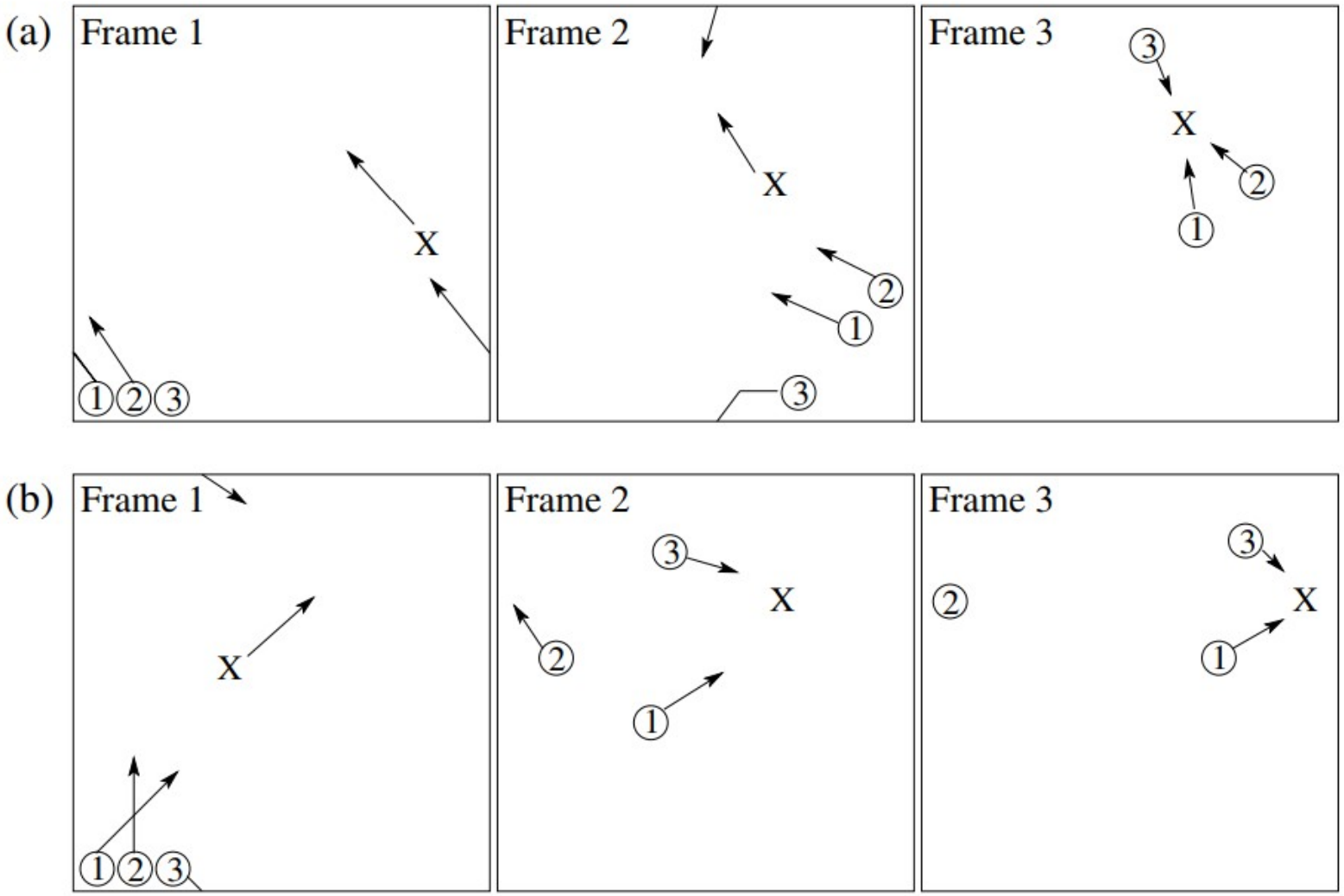


Fig. 11. **Two sample strategies of the same communicating team.** This team employs the two strategies shown above, as well as their variations and combinations. In the first, (a), the chase starts with two chasers and a blocker, but ends with opposite chasers. In the second, (b), there is a blocker and two chasers throughout, but the movement is horizontal. In this manner, the same team utilizes different strategies, depending on the starting position of the prey.

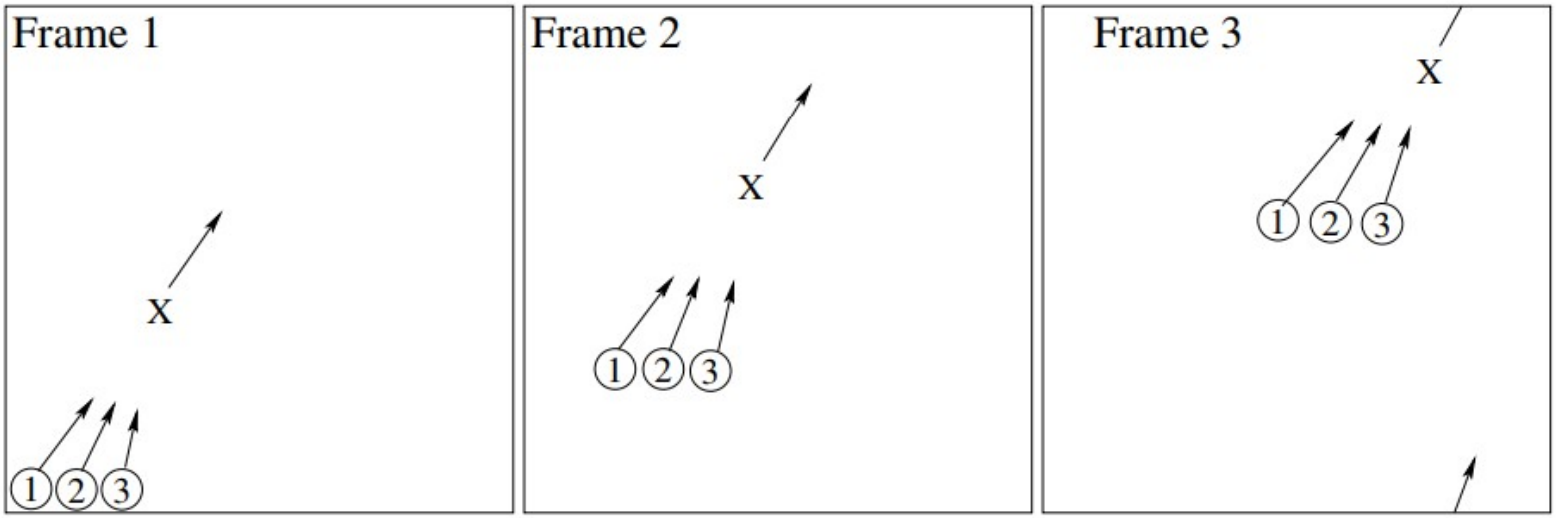


Fig. 18. **A strategy of three individually evolved predators placed on the same environment.** The predators chase the prey together in the nearest direction but are unable to catch it. Coevolution is thus essential in this task to evolve successful cooperative behavior.


VISIT www.OneLifeOnScreen.com