Predator–prey survival pressure

Rafaelle Lacraz, Titouan Steyer

Collective Behaviour Course, Research Seminar Report

November 2025

Supervisor: Iztok Lebar Bajec

This project explores how collective behaviors can emerge in predator–prey systems driven by minimal survival incentives. Following the framework of Li et al. [1], we implement a 2D continuous environment with multiple prey and a few predators, each governed by physical dynamics and subject to active and passive forces. The reward function is deliberately simple predators gain a reward when catching prey, and prey receive a penalty when caught. Using public code as a starting point, we reproduce baseline behavior and visualize system evolution under random motion. Quantitative metrics such as the Degree of Sparsity (DoS) and Degree of Alignment (DoA) allow us to characterize emergent group dynamics. This report serves as a progress checkpoint toward building an adaptive reinforcement learning model that captures realistic group behavior.

1 Introduction

Understanding how collective behaviours such as swarming, flocking, or coordinated pursuit emerge in animal groups is a central question in behavioural biology and bio-inspired robotics. Classical models rely on manually designed interaction rules, which can reproduce rich patterns but do not explain how such behaviours arise naturally through adaptation.

Li et al. [1] address this limitation with a minimalist predator–prey framework based on multi-agent reinforcement learning (MARL), where no social rules are explicitly encoded. Agents receive only a simple survival-based reward (+1 for a predator catching a prey, -1 for a prey being captured). Despite this simplicity, the system exhibits diverse emergent behaviours quantified by the Degree of Sparsity (DoS) and Degree of Alignment (DoA).

In this project, we reproduce and analyse this framework, starting from a simplified baseline environment. Our goal is to study how collective behaviours emerge, how learning influences predator–prey interactions, and how system parameters shape these dynamics.

2 Methods

2.1 Environment and Implementation

We chose to replicate the simulation environment described by Li et al. [1] to ensure consistent conditions for the emergence of collective behavior. This environment is a continuous two-dimensional square space (with side length $L = 2$ m) in which two types of agents interact: predators and prey. The environment supports two distinct types of boundary conditions:

Periodic boundaries (torus) The domain has no effective boundaries: when an agent exits from one side of the square, it reappears on the opposite side with the same velocity. These periodic edges connect the borders of the square, approximating an infinitely extended space.

Solid boundaries (walls) The domain is bounded by solid edges that agents cannot cross. The square edges act as walls with a repulsive force simulating contact (spring stiffness $k = 50$ $\text{N}\cdot\text{m}^{-1}$). When an agent reaches the boundary, it experiences a restoring force pushing it back inside the domain, preventing it from leaving the simulation space.

In our current implementation, we prioritize periodic boundaries (as a starting point identical to the paper), but we also plan to implement the solid-wall case later. In either case, the environment is initialized by placing a certain number of predators (n_0) and prey (n_1) at random positions with random initial orientations. Following the article, we start with a small population ($n_0 = 3$ predators and $n_1 = 10$ prey) for training, although these numbers can later be increased for testing. Agents of the same species are assumed to be homogeneous in capabilities.

Agent representation Each agent (predator or prey) is modeled as a disk, visually represented as a circle with a short line segment indicating its heading. All agents have mass $m = 1$ kg for simplicity. Predators and prey differ slightly in locomotor capacity: for instance, predator maximum speed is set to 0.35 m·s^{-1}, while prey is limited to 0.3 m·s^{-1}. Each agent's control is defined by two components (a_F, a_R) corresponding to forward propulsion force and rotational command, respectively. To ensure realism and stability in simulation, these actions are bounded: linear acceleration $|a_F|$ is limited to 1 m·s^{-2} and angular speed $|a_R|$ to 0.5 rad·s^{-1} (values taken from the original paper).

Active and passive forces According to the physical model from Li et al., agents are subject to both active (self-controlled) and passive (external) forces.

The active force results from self-propulsion. It has two components: (1) forward thrust aligned with the agent's heading direction h, scaled by a_F; (2) a torque that rotates the agent, with intensity a_R (bounded as above). These two components a_F and a_R are the outputs of the agent's controller at each time step.

The passive forces come from the simulated physical environment. These include: (1) viscous drag force f_d, which opposes motion and is proportional to velocity (drag coefficient set to 2 N·s·m^{-1}); (2) contact forces during collisions either between agents (f_a) or between agents and boundaries (f_b). These are modeled using Hooke's law (linear spring) with stiffness $k = 50$ N·m^{-1}, and they accumulate if an agent is in contact with multiple neighbors. This prevents agent overlap and creates realistic bouncing dynamics.

All these forces are updated at every discrete time step $\Delta t = 0.1$ s in our simulation. Each agent's motion is integrated numerically using Newton's second law. Our software implementation follows this physical model strictly, ensuring that the virtual environment closely reproduces the reference study.

2.2 Collective Behavior Metrics

In addition to visual inspection, we plan to use quantitative metrics to assess the emergent behaviors within the agent populations. Following the definitions introduced by Li et al. [1], we will compute two key indicators: the *Degree of Sparsity* (DoS) and the *Degree of Alignment* (DoA). These metrics are designed to evaluate local cohesion and heading synchronization within a species, while being invariant to global translations and rotations of the group.

Degree of Sparsity (DoS). The DoS $\in [0, 1]$ quantifies the spatial distribution of agents by averaging the normalized distance to the nearest neighbor among conspecifics. It is defined as:

$$\mathrm{DoS} = \frac{1}{TND} \sum_{t=1}^{T} \sum_{j=1}^{N} \left\| \mathbf{x}_j(t) - \mathbf{x}_{k(j)}(t) \right\|$$

where $\mathbf{x}_j(t) \in \mathbb{R}^2$ is the position of the j-th agent at time step t, and $k(j)$ is the index of its nearest neighbor: $k(j) = \arg\min_{\substack{k \in \{1,\dots,N\} \\ k \neq j}} \|\mathbf{x}_j(t) - \mathbf{x}_k(t)\|$ T is the episode length (number of time steps), N is the number of agents of the same type (e.g., prey), and D is the environment size, defined as the maximum possible distance between two agents. A high DoS value reflects a dispersed group, while a low DoS indicates strong cohesion and clustering among agents.

Degree of Alignment (DoA). The DoA $\in [0, 1]$ measures the similarity in heading directions between neighboring agents and is defined as:

$$\mathrm{DoA} = \frac{1}{2TN} \sum_{t=1}^{T} \sum_{j=1}^{N} \left\| \mathbf{h}_j(t) - \mathbf{h}_{k(j)}(t) \right\|$$

where $h_j(t) \in \mathbb{R}^2$ is the heading vector (unit norm) of agent j at time t, and $k(j)$ is again its nearest neighbor as defined above. The factor $1/2$ ensures that the metric remains bounded in $[0, 1]$.

A low DoA indicates poor alignment, while higher values reflect synchronized heading directions.

It is important to note that DoA is not equivalent to the mean heading of the entire group. In particular, a population may form multiple distinct flocks, each internally aligned but moving in different directions. In such cases, a global average heading would cancel out, but the local DoA would still be high—making it a more appropriate measure of flocking behavior.

These two metrics will be implemented in the next stage of our project and used systematically to evaluate behavioral patterns and the impact of learning policies.

2.3 Reward Function

We adopt a minimal reward structure, consistent with Li et al. [1], to promote emergent behaviors without artificial bias:

- Predators receive +1 when catching a prey.
- Prey receive −1 when caught.

No additional incentives (such as cohesion or alignment) are used. This choice allows agents to autonomously develop strategies based solely on their individual objective of survival or capture. We found a public GitHub repository [2] whose author based their work on the same article, and we used some of their Python files as a starting point for our own model implementation.

2.4 Baseline Objective

The goal of this first phase is to reproduce the baseline behavior described by Li et al. [1]:

- A continuous 2D environment with multiple prey and a few predators.
- Agents modeled as disks defined by their position, velocity, and heading.
- Implementation of a reward function based on predator-prey interactions.
- Visualization of the evolution of collective behavior metrics (DoS and DoA) and cumulative reward over time.

This baseline will serve as the foundation for further analysis of collective behaviour and for exploring modifications of the original model.

3 Results

3.1 Baseline observations without learning

As a first step, we visualized the interaction dynamics between three predators (in red) and multiple prey (in green) in the continuous 2D environment, before introducing any reinforcement learning mechanism. In this baseline setup, agents follow random motion policies — no social, cooperative, or strategic rules are explicitly defined. The environment uses periodic boundary conditions: when an agent crosses one side of the square domain, it reappears on the opposite side, ensuring an effectively infinite space without borders.

A short animation (Figure 1) illustrates a typical simulation sequence. Despite the absence of behavioural rules, random encounters and temporary clustering can occasionally be observed due to stochastic motion and spatial confinement.

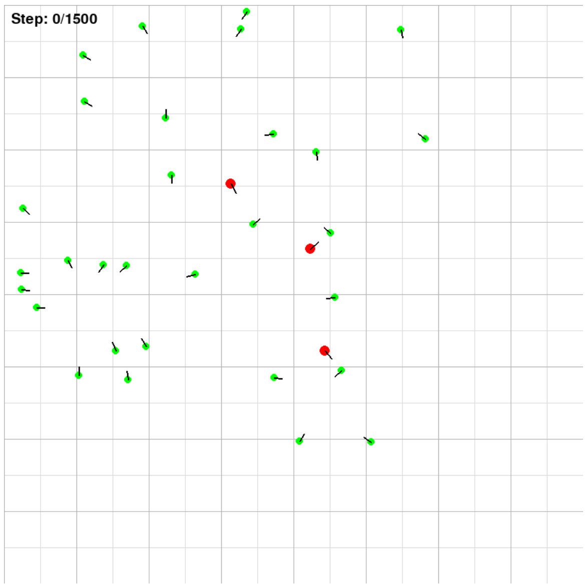

Figure 1: Baseline simulation with random motion: three predators (red) and multiple prey (green) evolving in a continuous environment with periodic boundaries.

To quantify these dynamics, we computed the Degree of Sparsity (DoS) and Degree of Alignment (DoA) throughout the simulation (Figure 2). In this random-motion setting, the DoS tends close to 1, indicating a uniform spatial distribution of agents, while the DoA remains near 0, confirming the absence of directional alignment. These values are consistent with the expected baseline behaviour reported by Li *et al.* [1], and validate that our environment behaves correctly before introducing learning mechanisms.

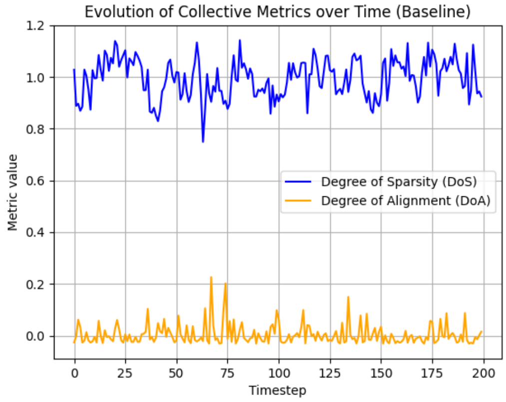

Figure 2: Evolution of the Degree of Sparsity (DoS) and Degree of Alignment (DoA) during the random-motion baseline simulation.

3.2 Extending the Model : Initial learning attempts

In a second phase, we aimed to increase the complexity of our model by introducing simple deterministic behaviours (predator pursuit and prey avoidance), and by implementing a reinforcement learning training procedure based on the Git repository we used as a reference.

The goal was to simplify an originally complex environment and obtain a workable version within a limited time frame. After several tests, the results remain only partially conclusive, and further improvements are still in progress.

Figure 3 shows the evolution of the predators' average reward throughout training. We observe a global upward trend, suggesting that the agents do start learning useful behaviours. However, the curve remains highly noisy, indicating that the model still requires refinement (parameter tuning, observation structure, reward shaping) before achieving stable and consistent learning dynamics.

4 Discussion

For the next milestone, our first objective is to stabilise the environment so that it accurately reproduces the behaviours described in the paper and yields the expected results. We will then focus

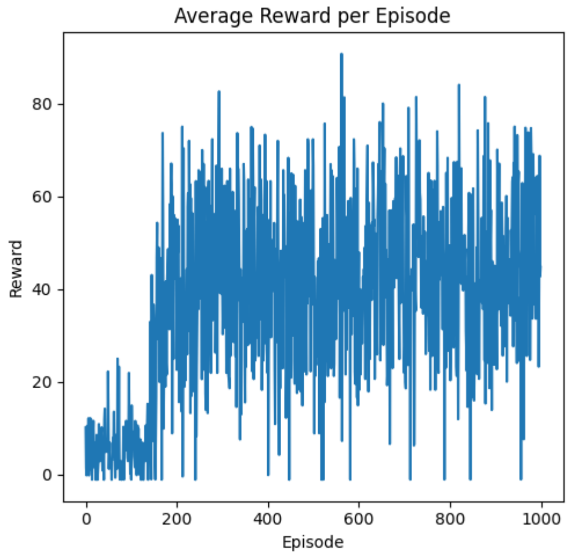

Figure 3: Average reward per episode during reinforcement learning

on improving the current training process, which is still incomplete and unstable. Once these two steps are achieved, we will be able to analyse emergent behaviours more rigorously using the metrics presented in the document.

From that point, we plan to adjust certain model parameters—such as the number of prey and predators or the observation radius—in order to study their influence on the system's dynamics.

If time permits, we also aim to explore the limits of the model by modifying more structural aspects: introducing faster prey or slower predators, adding obstacles to the environment, or investigating scalability by varying the total number of agents.

References

[1] Zhengyang Li, Emanuele Cuccoli, Marco Villani, and Andrea Cavagna. Emergence of collective behaviour in coevolving predator–prey systems. *New Journal of Physics*, 25(9):093028, 2023.

[2] xxnnnnn. Predatorprey_rl_reproduction. https://github.com/xxnnnnn/PredatorPrey_RL_Reproduction, 2024.