# Modeling Social Distancing with Reinforcement Learning

Nejc Ločičnik, Igor Nikolaj Sok, Leon Todorov, Andraž Zrimšek

***ABSTRACT: This study explores the use of reinforcement learning (RL) to model social distancing behaviors in artificial agents, aiming to minimize disease transmission in a two-dimensional environment. Inspired by natural behaviors observed in species like ants, agents exchange health information and adapt their movements to avoid infected individuals. We investigate how different reward structures, such as collision-based penalties and external tasks like wall-touching, influence agent behavior and the emergence of social distancing. Results show that agents trained with RL exhibited increased separation between healthy and infected agents, reducing infection spread. The findings suggest that RL can be a powerful tool for simulating disease dynamics and informing swarm robotics and public health strategies.***

## I. Introduction

The spread of infectious diseases presents a significant challenge to both human and animal populations, prompting the development of mechanisms to minimize transmission. Social distancing, an adaptive behavior where organisms avoid infected individuals, has been observed across various species, providing an evolutionary advantage in disease mitigation. During the COVID-19 pandemic, social distancing became a key public health strategy, prompting interest in how such behaviors might emerge in artificial agents.

Modeling disease transmission and social distancing behaviors in simulated environments can provide valuable insights for fields like epidemiology, robotics, and swarm intelligence. Traditional methods often rely on predefined rules, limiting the complexity and adaptability of agent behaviors. Reinforcement learning (RL), however, offers a flexible framework where agents learn to adapt based on reward structures, fostering more organic behaviors that evolve in response to environmental pressures.

In this study, we used RL to model social distancing behaviors in agents within a two-dimensional environment. By adapting their interactions based on health information exchanged with one another, agents aim to minimize disease transmission. Building on predator-prey dynamics in multi-agent RL frameworks, we explored how different reward structures and network adaptations contribute to the emergence of social distancing, where agents autonomously avoid infected individuals.

Our goal was to deepen the understanding of social distancing behaviors in adaptive multi-agent systems, with potential applications in disease transmission modeling, swarm robotics, and public health simulations.

## II. Related Work

In Predator–prey survival pressure is sufficient to evolve swarming behaviors [1], the authors used reinforcement learning (RL) to model predator-prey dynamics in a cooperative–competitive multi-agent framework. Predators are rewarded for catching prey, and prey for avoiding capture, allowing agents to develop adaptive behaviors such as swarming and predator dispersion. Unlike traditional rule-based models, RL enables emergent, flexible behaviors that better reflect real-world complexity. Drawing from this, we adapted RL methods to model disease spread, tuning agent parameters and rewards to simulate social distancing behaviors.

Social distancing as a natural response to disease is further explored in Infectious diseases and social distancing in nature [3], which examines how animals instinctively adjust interactions to reduce transmission. These behaviors arise through precautionary actions by healthy individuals or physiological responses in infected ones. Meanwhile, Romano et al. [2], in The trade-off between information and pathogen transmission in animal societies, highlight the balance between minimizing infection risk and maintaining critical social connections. They proposed "network plasticity" as a mechanism enabling populations to optimize this trade-off, offering insights into decision-making within social groups.

Together, these studies provide frameworks for modeling adaptive behaviors under environmental pressures. Leveraging these insights, we use RL to simulate the dynamics of disease transmission, focusing on how agent interactions influence social distancing and infection spread.

## III. Methodology

Our objective is to simulate the spread of infectious diseases within a population of agents navigating a two-dimensional environment, exploring how they can autonomously learn to minimize transmission by adapting their interactions based on health information exchanged with others.

### A. Simulation Environment

This study employed a multi-agent reinforcement learning (RL) framework, adapted from the environment developed by Li et al. (2023) [1]. The simulation takes place in a two-dimensional continuous space with periodic boundary conditions, meaning that agents crossing one edge of the square environment reappear on the opposite edge, retaining their velocity.

Agents were modeled to resemble ants (changed from unicycles like in Figure 1), with their body consisting of three connected circles (the back circle being slightly larger) and six legs. Their behavior is driven by a combination of active and passive forces. Active forces, controlled by the agents, include a forward movement force ($aF$) aligned with their heading direction and a rotational force ($aR$) enabling changes in heading. Passive forces, inherent to the environment, include drag force ($Fd$), simulating resistance opposing the agent's velocity, and repulsive force ($Fa$), which prevents agents from overlapping by pushing them apart. At each simulation step, the agents' positions and velocities are updated by summing all acting forces, with the dynamics governed by:

$$\dot{x} = v, \quad \dot{v} = \frac{ha_F + F_d + F_a}{m}, \quad \dot{\theta} = a_R$$

where $x$ is the agent's position, $v$ its velocity, $\theta$ its heading angle, $h$ the unit vector for heading direction, and $m$ the agent's mass. The agents along with the forces acting on them are illustrated in Figure 1.

To tailor the framework to our objectives, we implemented several modifications: 1) changed agent visualization from unicycles to ant-like figures and adjusted movement parameters; 2)

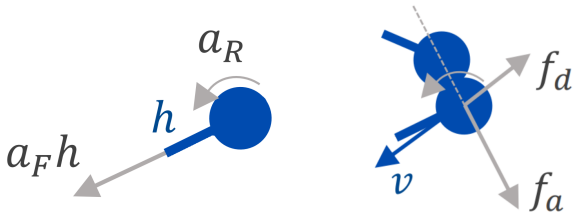


Figure 1. Active (left) and passive (right) agent forces. [1]

incorporated health status perception for each agent; 3) tracked agent interactions for network evaluation; and 4) redefined the reward policy to align with disease-spread mitigation goals.

### B. Basic Reward Policy

Our initial reward policy aimed to produce social distancing patterns in agent behavior is based on direct collisions between agents as the primary form of information exchange. Collisions between agents of the same status (healthy-healthy or infected-infected) were rewarded to encourage grouping behavior. In contrast, collisions between agents of different statuses (healthy-infected or vice versa) were penalized to discourage close contact to limit disease spread.

The above mention policy is demonstrated in isolation with all healthy agents in Figure 2. On the left simulation we penalized each agent upon collision with a $-1$ reward, while on the right simulation we rewarded each colliding agent with a $+1$ reward. This demonstrates how a very simple change in the reward policy effects the learned behavior.

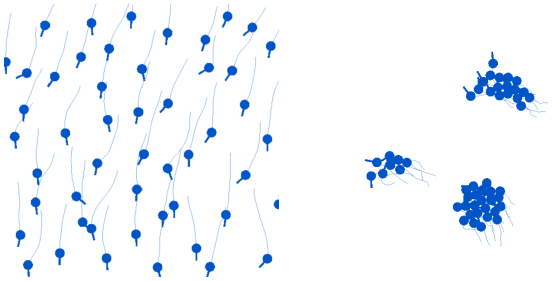

Figure 2. Demonstration of simple reward policies - penalize (left) or reward (right) collisions.

We tried a few approaches to augment the simple reward policy:

1) **Diminishing Reward for Interactions** $\rightarrow$ To reduce agent clustering, consecutive interactions were progressively penalized based on recent interactions ($h(x) \rightarrow$ health status of agent $x$).

$$\text{reward}(a,b) = \begin{cases} -\lambda & \text{if } h(a) \neq h(b) \\ +\sigma * \gamma(1 - recent(a,b)) & \text{otherwise.} \end{cases}$$

2) **Wall Collision Penalty** $\rightarrow$ In non-periodic environments, we penalized agents for colliding with walls.

$$\text{reward}(a) = \begin{cases} -\lambda & \text{if } a \text{ collides with any wall} \\ 0 & \text{otherwise.} \end{cases}$$

3) **Energy Control Penalty** $\rightarrow$ A control penalty was introduced to simulate energy consumption, penalizing the magnitude of the agent's control inputs.

$$\text{reward}(aF, aR) = -(\alpha|aF| + \beta|aR|)$$

However, we found these approaches too engineered for our goals and ultimately decided to stick with a simpler reward policy.

What we did find to vastly improve visualization results was the introduction of an external task. This task added complexity that better reflected real-world dynamics. Agents alternated between touching the left and right walls, with each agent initially assigned a wall. They were rewarded $+0.1$ for moving in the correct direction and $+1$ for reaching the assigned wall. After completing the task, agents switched to the opposite wall, requiring them to continuously adapt their movement strategy. This task significantly enhanced the agents' performance and provided clearer insights into their behavior.

### C. Alternative Information Exchange

Our main exploration of agent interaction (information exchange) was on the basis of collisions. As an alternative, we also considered a pheromone-based system inspired by ant behavior, where agents perceived "safety" or "danger" through accumulated, decaying concentrations. While promising in theory, this approach performed poorly compared to integrating movement-based tasks. The core issue was the lack of a clear relationship between pheromone observations and the actions needed to maximize rewards. This underscores the importance of designing observation mechanisms that directly guide actionable, reward-driven behavior.

## IV. Results

### A. Interaction Network Observations

To assess the emergence of social distancing behaviors, we convert agent interactions into a network as described in Stroeymeyt et al. (2018) [4]. During each evaluation step, interactions are recorded in an $n \times n$ matrix, where $n$ is the total number of agents. A collision between agents is considered an interaction. After each evaluation, the matrix is normalized to create the network, where nodes represent agents and edges form if interaction values exceed a 0.01 threshold. Edge weights correspond to interaction values, and nodes are labeled with the agents' health status.

We evaluated our trained model over a $10,000$-step episode using both a random, untrained network and a trained network, constructing interaction networks to analyze agent behavior. The random network lacked structure, forming a mostly fully connected graph. In contrast, the trained network exhibited separation, with infected agents interacting less with healthy ones while maintaining interactions with other infected agents. When edges were weighted by interaction frequency, these patterns became clearer, as shown in Figure 3.

We calculated network metrics such as clustering, modularity, and density to quantify agent interactions. Modularity increased from $-0.0005$ (random) to 0.25 (trained), and clustering rose from 0.056 to 0.066, indicating greater group separation and the emergence of social distancing. These changes align with reduced infection spread, as suggested in [5].

To study pathogen-induced changes, we ran 10 experiments using the best-performing model. In each, agents were designated as infected, and a baseline $10,000$-step evaluation was conducted without actual infections. Metrics were recorded before and after introducing infections. Results in Figure 4 show increased modularity as infected agents segregated and decreased network efficiency, consistent with reduced pathogen spread. However, clustering unexpectedly decreased due to the

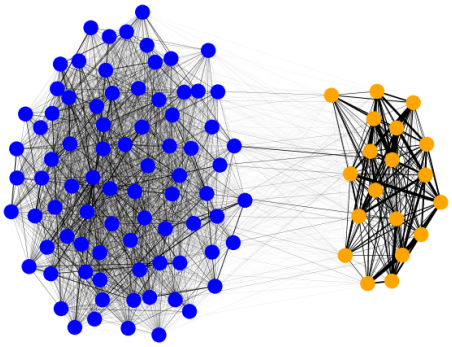

Figure 3. Filtered learned interaction network.

high connectivity of the initial infection-free network, where agents formed dense groups with a clustering coefficient of 0.138.

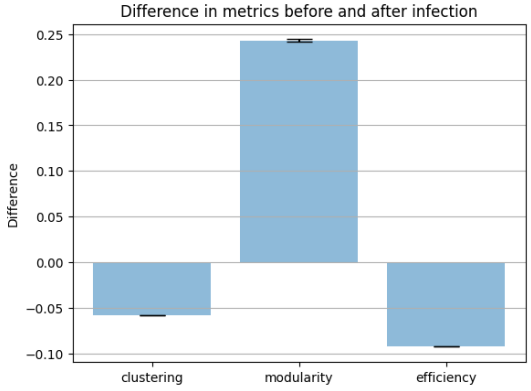


Figure 4. Changes in interaction network properties when introducing infected agents.

The results, shown in Figure 4, reveal a significant increase in modularity, as infected agents became more segregated from the healthy population. Additionally, network efficiency decreased, suggesting an increase in the shortest paths within the network, which is consistent with reduced pathogen spread. However, contrary to expectations, clustering decreased. This anomaly can be explained by the initial lack of infected agents, resulting in a highly connected network with a high clustering coefficient of 0.138 due to the agents' tendency to form dense groups.

### *B. Simulation Visualization Observations*

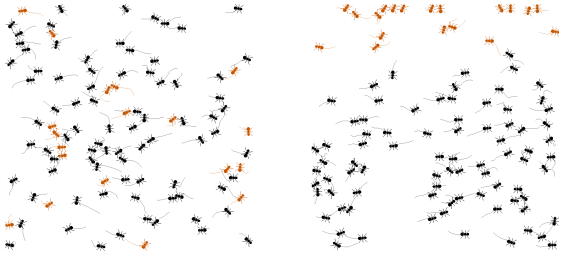

Figure 5. Environment simulation visualization: start (left) vs. end (right).

Figure 5 illustrates a simulation of our best-performing policy. Orange ants represent infected agents, while black ants represent healthy ones. The left image (frame 25) shows the start of the simulation, and the right image (frame 1300) depicts a later stage. A clear separation between the two groups is observed, minimizing contact between agents of different health statuses. Notably, this separation only emerged after introducing the additional left-to-right movement task. Without this task, while the interaction network properties remained similar, the visual separation was far less pronounced.

## V. Discussion

This study used reinforcement learning (RL) to model social distancing behaviors in artificial agents, aiming to minimize disease transmission in a two-dimensional environment. Inspired by natural behaviors like those of ants, agents exchanged health information and adapted their movements to avoid infected individuals, simulating social distancing.

Results showed that agents trained with a basic reward policy increased the separation between healthy and infected agents, as reflected in network metrics like modularity and clustering. These metrics indicated that agents successfully minimized disease transmission by avoiding contact with infected individuals.

We also examined the impact of reward components, including diminishing rewards, control penalties, and wall collision penalties, which refined agent behavior and encouraged more efficient movement patterns. Additionally, agents alternated between touching walls, promoting more structured behavior and facilitating social distancing. In contrast, a pheromone-based exchange system was less effective due to its misalignment with reward-driven actions.

Our findings suggest RL can effectively model disease spread and social distancing behaviors. Future research could build on these results by incorporating more complex disease dynamics, improving agent interactions, and applying the model to public health simulations and swarm robotics.

**CONTRIBUTIONS**: **LT** prepared/fixed the environment setup and did the basic avoid/touch experiments, **INS** and **AZ** did the reward policy experiments and the network statistics, **NL** did the alternative interactions experiment and organized/polished the report. Each member wrote their own parts of the report.

The source code for this project is accessible on the GitHub Repository.

## References

[1] Jianan Li, Liang Li, and Shiyu Zhao. "Predator–prey survival pressure is sufficient to evolve swarming behaviors". In: *New Journal of Physics* 25.9 (2023), p. 092001.

[2] Valéria Romano, Cédric Sueur, and Andrew JJ MacIntosh. "The tradeoff between information and pathogen transmission in animal societies". In: *Oikos* 2022.10 (2022), e08290.

[3] Sebastian Stockmaier et al. "Infectious diseases and social distancing in nature". In: *Science* 371.6533 (2021), eabc8881.

[4] Nathalie Stroeymeyt et al. "Social network plasticity decreases disease transmission in a eusocial insect". In: *Science* 362.6417 (2018), pp. 941–945.

[5] Erik M Volz et al. "Effects of heterogeneous and clustered contact patterns on infectious disease dynamics". In: *PLoS computational biology* 7.6 (2011), e1002042.