In this work, we analyze the impact of simple social signaling mechanisms in the performance of learning agents in competitive multiagent settings. The social signals influence the total amount of reward received throughout time.

In a series of experiments, we measure the combined performance of agent populations according to distinct selection paradigms.

The results of our study show that agents learn to collectively coordinate their feeding behavior by trading-off immediate benefit for long-term social welfare.

The individuals with the highest fitness appear in the most socially-aware populations.

Social inequalities arise where the more fit agents benefit from their higher status being appreciated by other members of the population.

Social Display
Signals allow for coordination by informing action opportunities to others, allowing groups of agents to perform better as a whole (e.g., cooperative hunting, gathering, breeding, etc.).

Repeated interactions between signaling individuals make each adjust its own strategy towards maximizing its payoff given some fitness function.

Competitive Multiagent Learning
Populations of $K$ agents, food exists for $K \times P_{self,max}$.

Partially-Observable Markov Game (POMG)
Each agent $k \in K$ is paired with another agent, $\ell$ according to $P_{self,max}(k, \ell)$

Partial state view (observation):
- Own satiation status, $U_k = \{\text{FULL}, \text{ALMOST_HUNGRY}, \text{HUNGRY}\}$
- Other’s status, $O_{k,\ell} = \{\text{NOTHING}, \text{OTHER_HUNGRY}, \text{OTHER_FULL}\}$

Select action according to current policy from $A_k = \{\text{Eat, Nothing}\}$

Results
Social Influence on Population Fitness
Positive impact on the population fitness. Agents achieve coordination, i.e. eat when necessary.

Impact of Available Resources
Avg. fitness: social signals important when half of populations is allowed to eat.
Max. fitness: most important when few resources are available
Appearance of “popular” or “high status” individuals.

Sensitivities
$
\theta_{ext} = \{0.0, 0.2, 0.4, 0.6, 0.8\}
$

Population reward function according to $P_{self,max}$.
Each agent is a self-interested, independent learner, each following and learning its own policy so to maximize the reward $r$.

Agent performance according to:

$$f_k = \sum \phi_{fit}(r)$$

Population Fitness
Measure the population performance, selection paradigms:

- Average fitness: $f_{avg} = \frac{1}{K} \sum_{k=1}^{K} f_k$
- Maximal fitness: $f_{max} = \max f_k$
- Std. dev. Fitness: $f_{std,dev} = -\sqrt{\frac{1}{K} \sum_{k=1}^{K} (f_k - f_{avg})^2}$

Experimental Procedure
Objectives
- Study complex dynamics arising from the agents actions.
- See conditions for the emergence of cooperation.
- How population fitness varies according to selection.

Sensitive analysis

- Experimental setup
  - $3300$ populations of $K = 100$ agents
  - $60$ independent Monte-Carlo trials
  - $1000$ time steps
  - Policy gradient algorithm