Human vs. Deep Neural Network Performance at a Leader Identification Task

Ankur Deka and Katia Sycara
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Phillip Walker
Smart Information Flow Technology
Minneapolis, MN 55401

Huao Li and Michael Lewis
School of Computing and Information
University of Pittsburgh
Pittsburgh, PA 15260

Control of robotic swarms through control over a leader(s) has become the dominant approach to supervisory control over these largely autonomous systems. Resilience in the face of attrition is one of the primary advantages attributed to swarms yet the presence of leader(s) makes them vulnerable to decapitation. Algorithms which allow a swarm to hide its leader are a promising solution. We present a novel approach in which neural networks, NNs, trained in a graph neural network, GNN, replace conventional controllers making them more amenable to training. Swarms and an adversary intent of finding the leader were trained and tested in 4 phases: 1-swarm to follow leader, 2-adversary to recognize leader, 3-swarm to hide leader from adversary, and 4-swarm and adversary compete to hide and recognize the leader. While the NN adversary was more successful in identifying leaders without deception, humans did better in conditions in which the swarm was trained to hide its leader from the NN adversary. The study illustrates difficulties likely to emerge in *arms races* between machine learners and the potential role humans may play in moderating them.

INTRODUCTION

As full scale deployment of robotic swarms for military missions as varied as logistics, surveillance, or combat nears, it is becoming increasingly important to devise ways to protect our own swarms and disrupt or destroy those of our adversaries. Key advantages of military swarms lie in minimizing readily detectable long distance communication and resilience to attrition and decapitation.

Swarms

Coordination among swarm members typically depends on some form of consensus algorithm by which swarm members exchange values, whether by observation or communication, and adjust their own parameters to reflect the local consensus. In flocking (Reynolds, 1987), for example, swarm members move away from others in close proximity, toward those at greater distances, and align their heading and velocity with those at a middle distance. This results in an emergent flocking behavior similar to that of birds or schooling fish (Couzin, et al., 2002) in which members move as a group albeit with continuing individual adjustments. Other emergent behaviors such as rendezvous (members converge on consensus location) or dispersion (members move away from one another producing an expanding perimeter) can be obtained by adjusting the attractive and repulsive forces.

The challenge to supervising/influencing such an autonomously coordinating swarm lies in biasing the consensus being computed locally across the swarm. Directly influencing all members by a mechanism such as broadcasting leads to more rapid convergence (Amraii et al., 2014) but at the cost of weakening consensus and potentially isolating members from the swarm. Influencing the swarm through leaders who contribute operator-biased parameter values to their neighbors provides a consensus preserving path of control that additionally benefits from a direct correspondence between input (heading and velocity in the case of flocking) and desired effect on the swarm. Herding (Long et al. 2020) offers a bio-inspired alternative to leaders based on sheep dog behavior and repulsion. In herding, however, the operator cannot conveniently influence the entire swarm and in particular, for multiple controlled herders, must account for complex interactions to accomplish a desired effect. In the third leading contender, potential fields, artificial potentials associated with fixed locations (Kolling et al., 2013) or virtual leaders (Kira & Potter, 2009), are used to attract/repel nearby swarm members thereby influencing consensus although again subject to complex interactions. Controlling a swarm through control of a leader(s) has, so far, been the most widely adopted strategy. Control via one or a small number of leaders has the advantage of limiting long distance communication while allowing the leaders to propagate commands within the swarm and synthesize information from the swarm to return to the commander while maintaining consensus and direct correspondence between inputs and desired outputs.

Leader Hiding

Resilience is one of the primary advantages of robotic swarms. While a multi-robot system (MRS) relying on hierarchical plans, repairs, etc. can execute far more complex plans, these plans may be brittle and disrupted by the loss of key players or capabilities. A swarm by contrast is fully robust to loss of robots as the remaining members continue to compute and follow their consensus. This crucial advantage, however, is jeopardized when control via leaders is introduced. The swarm is now subject to decapitation and, if leaders can be identified, may be even more vulnerable than a conventional MRS. While solutions such as dynamically designated leaders (Walker et al., 2014) have been proposed, redesigning swarm behavior to hide leader(s) may be the more straightforward answer.

To understand how hiding leader identity might be possible it is important to consider what is meant by a leader in the context of swarm supervision. The relation between the leader and other members of the swarm is informational. Swarm algorithms are agnostic as to whether parameter values are exchanged through communication or sensed as observations. Leader identification within a flock is difficult because the consensus algorithm controlling other members, camouflages the small deviations introduced by the leader(s). Larger deviations are infeasible because they would lead to loss of connectivity and dissolution of the swarm. The leader identification problem, therefore, becomes one of identifying the individual(s) that by some small amount, influence their fellows to a greater extent than their fellows influence them. An observer finds a leader by searching for a swarm member whose behavior cannot be completely explained by an estimate of its prior state and the values and changes among its neighbors. As this suggests, the leader identification decision is difficult and for a human would depend on complex judgments based on factors such as the Gestalt principle of Common Fate (Sturzel & Spillman, 2004).

Zheng et al. (2020) published the first evaluation of a leader hiding algorithm. They started with a swarm following the simple Reynolds (1987) flocking algorithm described above in which a leader led the swarm through a sequence of waypoints. Zheng et al. (2020) developed what they called privacy preserving flocking using an alternating optimization procedure with genetic algorithms in which controller parameters were optimized for preserving privacy during flocking optimization, while leader identification accuracy was optimized in the following discrimination learning phase. While the discriminator, termed the adversary, grew increasingly more powerful over the course of their experiment the privacy preserving flocking algorithm also improved and was particularly effective in preserving privacy from the discriminator over curvilinear paths.

In this paper we describe the development of private flocking using multi-agent reinforcement learning (MARL) in an adversarial framework similar to that of Zheng et al. (2020). Flocking swarm members and their adversary were trained in phases: first the swarm to follow the leader to the goal, then the adversary to identify that leader, and finally the swarm to hide its leader from a learning adversary. Human participants were then tested under the same conditions as the adversary to assess their relative performance in detecting the identity of hidden leaders.

METHODS

Adversarial Training

Adversarial training pitting two optimizers against one another as Zheng et al. (2020) is a standard training method. In the case of a literal adversary such as a military opponent attempting to ward off an attacking swarm, however, the symmetry in training is broken. Those developing the swarm are motivated to train it to avoid leader detection while withholding training trajectories from the adversary. The adversary on the other hand is motivated to obtain updated trajectories in order to counter these hiding techniques. In the resulting arms race the leader hiding swarm and the adversary attempting to identify its leader are actual adversaries. This competitive process between state actors can be simulated by conventional adversarial training which is conducted in phases such that the adversarial discriminator does not have access to the swarm's training trajectories prior to the onset of hostilities but subsequently can train in tandem from observations to improve leader identification. The correspondence between adversarial training and the expected course of development of leader hiding and refinement of detection countermeasures is used in this study to explore potential hazards of weapon systems controlled by learning algorithms and potential roles for human supervisors in this increasingly automated process.

Multi-agent Reinforcement Learning (MARL)

While conventional swarms employ controllers using local control laws, in this study we replace controllers with neural networks performing the same function in order to provide a framework more conducive to learning to hide the leader. This formulation is task agnostic, allows for additional consideration of adversarial component, and eliminates the requirement of hand-crafted swarm controllers, tuning of parameters, or expert knowledge for any given primary objective of the multi-agent system. MARL algorithms that use centralized training with decentralized execution are used for our problem as it is reasonable to remove computational and communication restrictions at training time and these algorithms still let all the agents act independently at test time (Lowe et al., 2017, Rashid et al., 2020). We use Graph Neural Networks (GNNs) to model the multi-agent architecture with agents modeled as nodes in a graph and inter-agent interactions modeled by the graph edges (Deka & Sycara, 2020). We extend the GNN framework to incorporate the leader-follower constraints of our problem. Our privacy-aware MARL framework follows the paradigm of centralized training with decentralized execution. This means that each agent maintains a separate copy of their respective parameters at test time so that it can perform its computations independently. The number of learn-able parameters in both the swarm and the adversary are independent of the number of agents or the time duration. As a result, these models are adaptable both temporally and in the number of agents. These are appealing properties for real world applications, e.g., depending on the task at hand we might wish to use fewer or more agents for the mission or some of the robot agents might get damaged during the mission. Swarm and adversary were trained using a multi-stage training process incorporating on-policy model-free reinforcement learning and supervised learning to train the swarm and the adversary respectively.

Training Phases

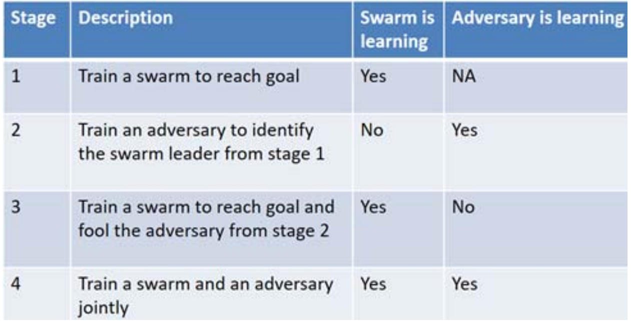
Swarms and Adversaries were trained and tested in a four phase sequence shown in table 1.

Phase 1: In the first phase swarms were trained in their novel MARL architecture to successfully follow a leader to a goal. Without conventional controllers, swarm members needed to learn from scratch relations between one another and their leader based on rewards obtained from reaching the goal. While the resultant behavior is not an exact replication of leader based on rewards obtained from reaching the goal. While the resultant behavior is not an exact replication of conventional (Reynolds, 1987) flocking, it reflects the distinction between a leader with knowledge of goal and followers who must base their behavior on observable aspects of the leader's behavior. The swarm at this phase corresponds to one trained for a task but not concealment.

Phase 2: In the second phase an adversary was trained to identify the leader for swarms trained in Phase 1. Swarm behavior was fixed and the adversary allowed to train to recognize leaders until an asymptote was reached. The adversary at this phase corresponds to a NN trained, in isolation from an opponent, to identify a leader for which no effort has been made for concealment.

[1]

Table 1. Training Phases



Stage	Description	Swarm is learning	Adversary is learning
1	Train a swarm to reach goal	Yes	NA
2	Train an adversary to identify the swarm leader from stage 1	No	Yes
3	Train a swarm to reach goal and fool the adversary from stage 2	Yes	No
4	Train a swarm and an adversary jointly	Yes	Yes

Phase 3: In the third phase the naïve adversary was fixed and the swarm trained to hide its leader until no further improvement was obtainable. The swarm at this phase corresponds to a swarm trained in isolation from its opponent to conceal its leader, prior to exposure in an observable conflict.

Phase 4: In the fourth phase the swarm and adversary were jointly trained in alternating trials to detect and conceal the leader. Performance at this phase reflects an arms race after both sides reveal observable data in which concealment and detection co-evolve. Because of the sensitivity of NNs to hyperparameters we report 3 of the most favorable selections.

Human participants were later asked to identify leaders for the same trajectories judged by the adversaries to provide a direct comparison between NN and human leader identification performance under adversarial conditions.

Participants

We recruited 81 participants from Amazon Mechanical Turk (Buhrmester, Kwang & Gosling, 2011) who were paid 1 USD for participating in a 30-minute online study. Data from one participant was lost. The remaining participants were randomly assigned to three groups to view swarm trajectories from either Phase 2 (24 participants), 3 (27 participants) , or 4 (4.1-12, 4.2-8, and 4.3-9 participants).

Simulation

Swarm trajectories were visualized in the Multi-Agent Particle Environment, MAPE (Lowe et al. 2017), a light weight multiagent simulation environment made available through OpenAI, a nonprofit organization providing tools and environments promoting the development of Artificial Intelligence including the widely used OpenAI Gym. MAPE provides a simple simulated physics in a discrete action, continuous observation space. Leader hiding methods described in this paper are currently being ported to cuSwarm (Walker, 2017), a larger scale parallelized swarm simulator developed to support research in human-swarm interaction.

Experimental Task

The NN adversaries and human participants observe the trajectory of a swarm portrayed as a collection of moving circless (figure 1) to the human and a sequence of (x,y) locations to the adversary as the simulated robots move from locations near the bottom of the screen to a randomly positioned goal near the top of the screen. Trajectory times ranged between 10-13 seconds.

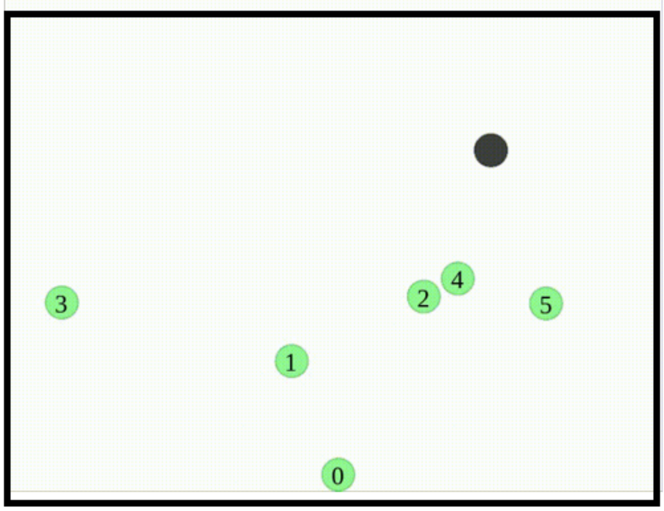


Figure 1 Display-goal is darkened circle

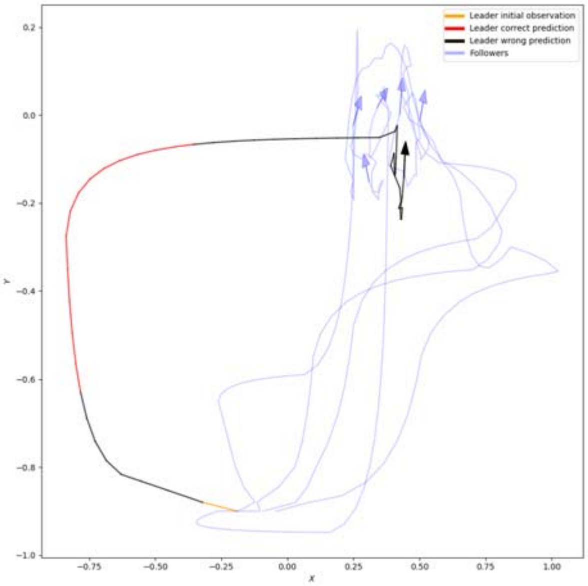


Figure 2 Leader correctly identified only for red portion of trajectory

The adversary made continuously revised estimates of the probability that each of the robots was the leader. The first four of these estimates were recorded at 2 second intervals while the final estimate was recorded at the conclusion of the trial. At any point the robot having the highest probability was considered to be the adversary's choice for leader. Figure

[1] We encourage the reader to try to identify the leader through an interactive web interface available at http://34.86.108.119/ (requires mouse and keyboard, tested on Chrome and Firefox browsers). This will not only show how we collected the data but also make the submission interactive.

2 shows a typical trajectory from Phase 4 in which the adversary misidentifies the leader at the start, makes correct identifications during the middle (shown in red), then lapses back into misidentification near the end of the trajectory. For purposes of analysis we identify the correctness of the adversary's identification to be the proportion of the trajectory for which it made a correct identification.

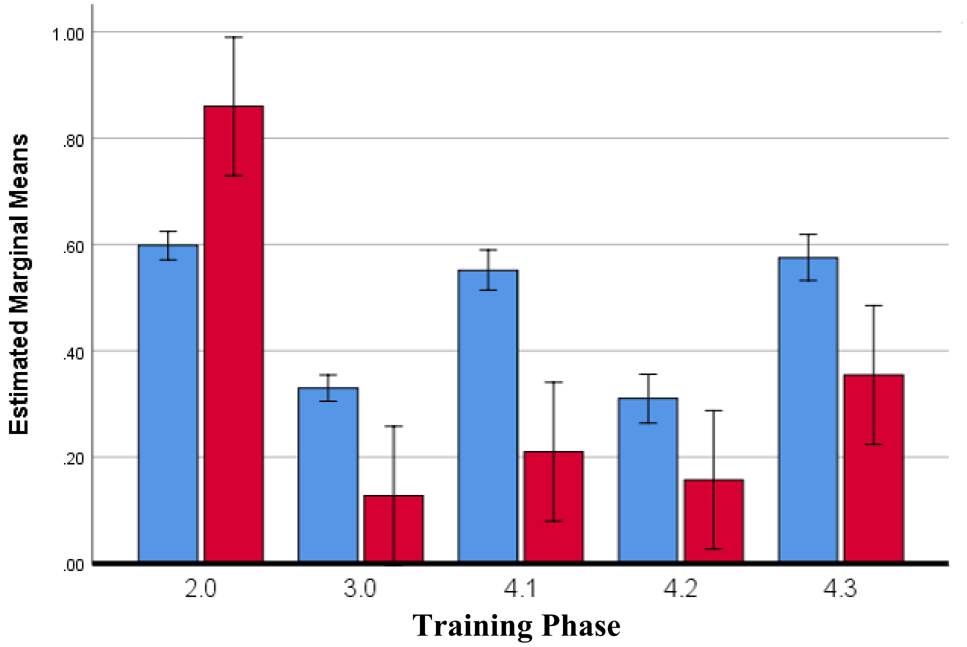


Figure 3. Percent correct identification, NN Adversary-red, Human Observer-blue

Human participants were instructed to push a button to stop the simulation as soon as they were able to identify the robot.they believed to be the leader from a menu, rated their confidence using a slider, then submitted their selection. Performance feedback was provided on the following screen from which the participant pressed a button to begin the next trial.

Swarms were fully trained before testing in each of the phases. For Phase 2 they were rewarded only for the navigation task in Phase 1. For Phase 3 they were rewarded both for the navigation task (going to the goal) and for hiding the leader (deceiving the adversary). In Phase 4 they were again rewarded for joint objectives but required to continue to evolve to avoid revealing their leader to an improving adversary. Human participants, after completing 3 practice trials, judged 50 trajectories from a single phase that were also used for testing the corresponding NN adversary.

Leader identification for humans was computed as the average of the correct identifications within their Phase group for comparisons between groups and with the adversary. Correlations of correct judgements between humans and the NN adversary were based on average value of human judgments and that of the adversary for each of the 50 trajectories.

RESULTS

A two-way ANOVA found main effects for correct identifications for phases ($F_{4,4249}$ =36.255, p < .001), between human and NN adversaries ($F_{4,4249}$ =36.255, p < .001) and for their interaction (($F_{4,4249}$ =11.361, p < .001). As shown in Figure 3 both human and NN adversaries did best in identifying leaders in Phase 2 where there was no training in concealment, worst in Phase 3 where swarm was trained to conceal without allowing training of adversaries, and intermediate in Phase 4 where both swarm and adversaries were allowed to train. Reported post hoc comparisons all employ Scheffè adjustment. All post hoc comparisons between phases were significant at the p < .001 level except Phase 4.1 which did not differ significantly from phases 3 or 4.3 which also did not differ significantly from Phase 2.

These differences were reflected in the effects of phase on human adversaries ($F_{4,3999}$ =70.422, p < .001) with post hoc comparisons showing differences at the p <.001 level between Phase 2 (M=.60), Phase 4.1 (M=.55) and Phase 4.3 (M=.58) with phases 3 (M=.33) and 4.2 (M=.31). The effects of adversary type and the interaction between adversary and phase reflect a prominent pattern in the results in which the NN adversary (M=.86) outperformed humans (M=.598) in Phase 2 ($F_{1,1249}$ =14.187, p < .001) for swarms without concealment but underperformed human adversaries in phases where concealment was present: Phase 3 (M=.127, M=.330, $F_{1,1399}$ = 9.203, p = .002), Phase 4.1 (M=.21, M=.552, $F_{1,649}$ = 23.16, p < .001), Phase 4.2 (M=.157, M=.293, $F_{1,449}$ = 5.338, p = .021), and Phase 4.3 (M=.355, M= .576, $F_{1,499}$ = 9.694, p = .002).

Effects favoring Phase 2 were found for confidence judgments ($F_{1,3999}$ = 28.908, p < .001) with all post hoc comparisons other than that between phases 3 and 4.1 significant at p < .023 or lower levels. Loss of confidence in identification of the adversarially trained leaders of Phase 4 were 3 (Phase 4.2) to 4 (Phase 4.3) times lower than the loss of confidence between Phase 2 and 3. A minor effect was found for latency ($F_{1,3999}$ = 4.214, p = .002), however, the only significant differences between phases were found between 4.2 with 4.1 and 4.3. We collected confidence and latency data in support of an evidence accumulation model (Evans & Wagenmakers, 2019) according to which later decisions should have greater confidence and greater accuracy due to the accumulation of evidence. Our data do not support this model as we did not find a significant correlation between confidence and latency. There were, however, small correlations between confidence (r=.119, p < .001) and latency (r= -.11, p < .001) with correct identifications.

To consider overlap in judgment between human and NN adversaries we examined the correlation of correct identifications between them on trajectories each had judged, finding an overall correlation of r=.253 (p < .001). The correlations within phases, however, tell a different story. In Phase 2 (r=.419, p = .002) and Phase 3 (r=.444, p=.001) where the NN adversary was not allowed to learn, the correlation between NN and humans was relatively high. In the three instances from Phase 4 where the adversary learned in tandem with the swarm the correlations were either nominally negative Phase 4.2 (r= -.091), Phase 4.3 (r= -.018) or significantly so Phase 4.1 (r= -.355, p= .011).

DISCUSSION

NNs are commonly perceived to be substantially more accurate than humans at difficult classification tasks such as our leader identification task. This is borne out by results for Phase 2 where there is no attempt at concealment. In the phases where concealment was employed, however, human observers delivered consistently superior performance. This should come as no surprise in that the adversarially trained swarm was explicitly trained to deceive the NN adversary and not a human observer. Vulnerability of NNs to adversarial noise, edge cases which are misclassified by the NN while readily apparent to a human, is an active research area in machine learning. A canonical problem from driverless car research involves the unrecognized stop sign in which small adversarial changes in appearance, usually determined from the NN's loss function, can be shown to disrupt recognition. A recent study by Evtimov et al. (2019) takes this phenomenon out of the lab by demonstrating that slight strategic alterations such as placing small advertising stickers on a physical stop sign can replicate this effect for real stimuli. Anomalies in learning such as biased training in which a classifier learns to distinguish wolves from huskies based on snow in the background (Ribeiro, Singh & Guestrin, 2016) are another common problem.

As critical functions in medicine, finance, and security come to rely on NN technologies it is especially important to understand and resolve such errors. For functions such as cyber security or military defense in which there is an inherently adversarial relationship, the problem is compounded because both parties have incentives to continue to learn in order to gain the upper hand. The machine learning arms race modeled in Phase 4 of our study provides a glimpse of what is likely to come. Some form of human-NN cooperation appears inevitable in attempting to meet this challenge. In phases 2 and 3 where learning was halted humans and the NN adversary made highly correlated judgments suggesting reliance on similar features. In adversarial Phase 4, by contrast, this correlation vanishes suggesting that when NNs were pitted against one another, humans and NN came to identify leaders in qualitatively different ways. Finding methods to exploit this divergence to produce synthesized classifications more accurate than either individually is a promising research approach.

ACKNOWLEDGEMENTS

This work was supported by ONR award N00014-19-C-1070 and AFOSR/AFRL award FA9550-18-1-0251.

REFERENCES

Amirpour Amraii, S., Walker, P., Lewis, M., Chakraborty, N. & Sycara, K. (2014) Explicit vs. Tacit Leadership in Influencing the Behavior of Swarms, *IEEE International Conference on Robotics and Automation (ICRA 2014),* 2209-2214.

M. Buhrmester, M., Kwang, T. & Gosling, T. (2011). Amazons Mechanical Turk:A New Source of Inexpensive, Yet High-Quality, Data? Perspectiveson Psychological Science, 6(1), pp 35.

Couzin, I., Krause, J., James, R., Ruxton, G. & Franks, N. (2002). Collective memory and spatial sorting in animal groups, Journal of Theoretical Biology, vol. 218, no. 1, pp. 1–11.

Deka, Ankur & Sycara, Katia (2021). Natural Emergence of Heterogenous Strategies in Artificially Intelligent Competitive Teams, arXiv preprint arXiv:2007.03102

Evans, N. J., & Wagenmakers, E. (2019). Evidence Accumulation Models: Current Limitations and Future Directions. PsyArXiv, https://doi.org/10.31234/osf.io/74df9

Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., Rahmati, A. & Song, D. (2018). Robust Physical-World Attacks on Machine Learning Models, IEEE Conference on Computer Vision and Pattern Recognition.

Kira, Z & Potter, M. (2009). Exerting human control over decentralized robot swarms, in 4th International Conference on Autonomous Robots and Agents, ICARA 2009. IEEE, pp. 566–571.

Kolling, A., Sycara, K., Nunnally, S., & Lewis, M. (2013). Human swarm interaction: An experimental study of two types of interaction with foraging swarms, Journal of Human-Robot Interaction, 2(2), 103-128.

Long, N., Sammut, K., Sgarioto, D., Garratt, M. & Abbass, H. (2020). A Comprehensive Review of Shepherding as a Bio-Inspired Swarm-Robotics Guidance Approach. IEEE Transactions on Emerging Topics in Computational Intelligence. PP. 1-15.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. Proceedings of the 30th International Conference on Neural Information Processing Systems NIPS.

Rashid, T., Samvelyan, M., de Witt, C., Farquhar, G., Foerster, J. & Whiteson, S. (2020). Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning, Journal of Machine Learning Research 21, pp. 1-51

Reynolds, C. (1987). Flocks, herds and schools: A distributed behavioral model. ACM SIGGRAPH Computer Graphics. 21. pp. 25–34.

Ribeiro, M., Singh, S. & Guestrin, C. (2016). Why Should I Trust You?": Explaining the Predictions of Any Classifier, Conference on Knowledge Discovery and Data Mining (KDD).

Sturzel, F. & Spillmann, L. (2004). Perceptual limits of common fate, Vision Research, vol. 44, no. 13, pp. 1565–1573.

Walker, P., Amirpour Amraii, S., Chakraborty, N., Lewis, M. & Sycara, K. (2014). Human control of robot swarms with dynamic leaders. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014): 1108-1113

Walker, P. (2017). Improving Operator Recognition and Prediction of Emergent Swarm Behaviors, Ph,D, dissertation, School of Information Sciences, University of Pittsburgh.

Zheng, H., Panerati, J., Beltrame, G. & Prorok, A. (2020). An Adversarial Approach to Private Flocking in Mobile Robot Teams. IEEE Robotics and Automation Letters 5, 2 (2020), 1009–1016