

On Leadership and Influence in Human-Swarm Interaction

Michael A. Goodrich, Sean Kerman, & Shin-Young Jung

Computer Science Department
Brigham Young University
Provo, Utah 84602

Abstract

In this position paper, we synthesize “within the system” models of human influence over bio-inspired swarms, summarizing observations from previous experiments and identifying methods of influence that have not yet been explored. We describe (a) differences among agents that can be controlled by a human and those that can't, (b) agents that are aware of the type of other agents and those that aren't, and the effects of attraction, repulsion, and orientation on human-guided swarm behavior. We also briefly discuss the interaction effort required to manage swarms.

Introduction

The relation between humans and swarms is often characterized in extreme ways. For control theorists seeking to design a desired set of emergent behaviors, the human is frequently considered a disturbance. They consider their “job” to design the system so robustly that no matter what wacky input some operator might inject, the system remains conservative enough to stay within the bounds of safety. While designing to this criterion can preserve the coherence of the swarm and guarantee certain properties such as convergence to consensus, it may so attenuate the operator's influence that the swarm becomes incredibly sluggish and continues to behave autonomously even as the human seeks to control it.

At the other extreme, ceding full authority to a human operator can likewise cause undesirable results such as (a) accelerating members at rates that lead to loss of coherence and break-up of the swarm, (b) highly inefficient state transitions that might be achieved much more smoothly if done in consonance with ongoing behaviors and (c) the possibility that the human is less informed than the agents. Moreover, centralized human control is susceptible to brittleness and sensitivity to communication delays or dropouts.

One possible way to mediate between these extremes is to use some sort of switching criterion wherein agents are responsible for behavior under some conditions and the human is responsible under other conditions. This recalls Sheridan and Verplank's levels of autonomy wherein responsibility for behavior is governed by predefined rules and conditions (Sheridan and Verplank 1978). The limitations of this

approach include (a) that the swarm is treated as a single unit that must coherently cede control to human input and (b) human input, when needed, asserts centralized authority over all agents. Such switching may be possible, but for swarms there is a powerful alternative.

We can find this alternative by acknowledging that a swarm, particularly under conditions of noise/error and limited bandwidth, may have more and deeper knowledge of its situation than a remote human operator. This knowledge is, of course, distributed among team members, and this distribution may make it difficult for a human to perfectly understand what the collective is doing. The human operator, however, has a better knowledge of mission goals but not necessarily of how the swarm may be able to achieve them. A solution is to require the operator to “work within the system” by injecting control through a small number of agents and allowing the system to adjust to these inputs over time. The resulting interaction uses the system's inertia to protect it from disturbance-like control “jolts” while remaining responsive to operator goals.

Thus, control authority manifests itself in two forms: persistence and span of control. Persistent influence requires that, to bring the system to a new course, the operator must persist in issuing a command over an extended time while persistently monitoring state information that may be noisy, delayed, or infrequent. Span of control is used by allowing the human some authority over the number of agents that he or she can influence.

Swarm Structures

There are many different types of swarm dynamics, but we will adopt the characteristics described below. The characteristics produce a large variety of swarm behaviors and structures and are fairly general.

Agent Dynamics

We assume that the agents can experience three different types of forces: repulsion, attraction, and orientation. These three fundamental forces have been used in a variety of agent models (Couzin et al. 2002; Spears et al. 2005; Reynolds 1987). These forces have been organized by Couzin to determine how agents respond to each other: agents that are with each other's zone of repulsion repel each other; agents that are within a zone of orientation align with each other;

and agents that are with a zone of attraction move close to each others (Couzin et al. 2002).

These zones, along with the notations used for the zones, are illustrated in Figure 1. Couzin’s model is thus a nonlinear controller where agents make zone-dependent switches between control laws.

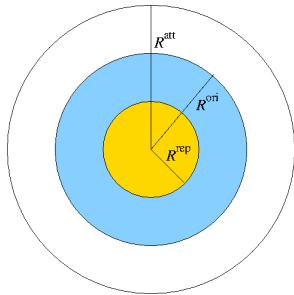


Figure 1: Couzin’s zones of repulsion, R^{rep} , orientation, R^{ori} , and attraction, R^{att} .

Phases

Quoting from (Goodrich et al. 2012), “[Figure 2] illustrates the four phases identified by Couzin. The figures illustrate agents as circles with a line emanating from them in their direction of travel. Each subfigure in [Figure 2] is a snapshot at one time instant. The *swarm phase* is a highly dynamic

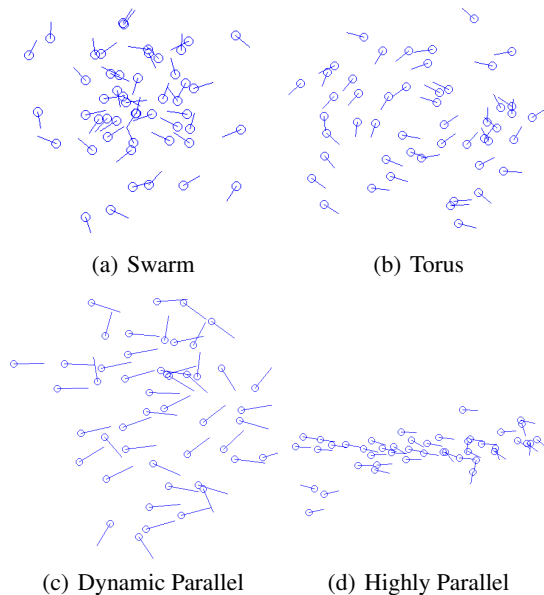


Figure 2: Snapshots of various phases from Couzin’s model: (a) Swarm. (b) Torus. (c) Dynamic parallel group. (d) Highly parallel group.

collective structure that tends to stay stationary and exhibits short bursts of agent alignment interspersed with apparently random milling about the swarms center. The *toroid phase*

has agents that circle in the same direction around a relatively stationary ‘hole’ in the middle of the structure. The *dynamic and highly parallel* structures both exhibit collective movement in a unified direction, with the highly parallel structure exhibiting greater alignment and greater group velocity.” We find it useful to lump the parallel phases together and call them “flocks”, and to mostly ignore the swarm phase. Instead, we will use the word swarm to include the torus and flock phases.

Note that, although this paper focuses on Couzin-like models, similar structures and similar results can be demonstrated for physicomimetic models (Goodrich et al. 2011b; Spears et al. 2005) suggesting that at least some of the results in this paper are likely to generalize.

Humans Influencing Swarms

In a series of experiments, we have investigated several approaches in which a human can influence a swarm without resorting to centralized control. It is helpful to make these forms of influence precise, and we choose to do so by presenting a simple mathematical formulation of the problem. In prior work (Goodrich et al. 2012; 2011a), we have presented similar formulations to the one used in this paper, but in this paper we use the formulation to identify other possible methods of influence. The formulation is not used to prove mathematically any properties of the system, but rather as a means for systematically identifying how a human can influence a swarm given clear assumptions.

Under the influence of a human, agent behavior can be responsive to three different stimuli: other agents in the collective, operator input, and signals from the environment. Restricting attention to an additive combination of these stimuli yields the following discrete-time model:

$$x_{t+1}^i = f(x_t^i, \mathbf{x}_t^{-i}; \theta^i, \theta^{-i}) + g(x_t^i, u_t^{\text{op}}; \theta^i) + e(\mathbf{x}_t^i, u_t^{\text{env}}; \theta^i). \quad (1)$$

In the model, x_t^i represents agent i ’s state at time t , \mathbf{x}_t^{-i} represents the states of every other agent in the collective except for agent i , θ^i denotes the agent’s type (described below), θ^{-i} denotes the types of every other agent in the collective except for agent i , f encodes how the states of other agents affect agent i , g encodes how human input, u_t^{op} , influences an agent with type θ^i , and e encodes how a signal from the environment, u_t^{env} influences an agent of type θ^i . Note that we have tried to keep this model as simple as possible by making an agent’s response to human input only a function of the agent’s state, x_t^i and not other states, \mathbf{x}_t^{-i} .

In this section, we will restrict attention to agents that can respond to human input but ignore environment input ($e(\cdot) = 0$). Such agents tend to produce the swarm-like structures described above, and adding a response to signals from the environment tends to create a different set of collective structures. We speculate about $e(\cdot) \neq 0$ later in the paper.

Agent Types

The agent’s type determines how it responds to human input, to external input, and to other agents. For any two types of agents θ and ϑ , $g^i(x_t^i, u_t^{\text{op}}; \theta) \neq g^i(x_t^i, u_t^{\text{op}}; \vartheta)$ meaning that

the agent types determine how an agent responds to signals from the human, with each agent type responding different than every other agent type.

We have explored two broad classes of agents: those that can respond to human input ($g() \neq 0$) and those that ignore human input ($g() = 0$). We call the first class of agents *human aware* and the second class *human blind*.

We have further explored two types of human aware agents. The first type of human aware agent is what we call a *special* agent. These agents are influenced only by the human and not by other agents, that is, $f = 0$ and $g \neq 0$. The second type of human aware agent is what we call a *stakeholder* agent. These agents are influenced by both the human and by other agents, that is, $f() \neq 0$ and $g() \neq 0$.

Among the human blind agents, we have two types: type aware and type blind. Since $g = 0$ for human blind agents, these agents cannot be directly influenced by a human. Instead, a human must influence these agents indirectly – by influencing an agent that then influences a reactionary agent.

A type aware agent can either process the states of other human blind agents or the states of nearby human aware agents, but not both. If a special agent or stakeholder agent is nearby, then the type aware agent is influenced only by that agent, but if no special or stakeholder agent is nearby, then the type aware agent is influenced only by other type aware agents. Thus, a type-aware agent is a nonlinear controller in which $f()$ depends on the types of the other agents with which it associates. The way a type-aware agent responds to the *states* of other agents depends on the *types* of those agents. For example, if a predator is close to a type-aware agent, then the type-aware agent ignores all other agents and flees from the predator, but if no predator is close, then the type aware agent switches and tries to align to its neighbors.

A type blind agent ignores human input since it is a human blind agent and it reacts to all agents within its zone of influence in the same way. It makes no distinctions between the types of the other agents, treating stakeholders, special agents, and other human blind agents the same way.

Influence Styles

In bio-inspired robot teams, two leadership models have been studied by others: lead-by-attraction and lead-by-repulsion (Olfati-Saber 2006). Lead-by-repulsion is commonly referred to as predation where the leader is a predator and agents are prey, so the leader influences the behavior of the agents by pursuing them. By contrast, lead-by-attraction is often associated with the colloquial use of the word leadership, meaning that a leader is one that gets ahead of a group and the group follows.

Given that inter-agent influence allows attraction, repulsion, and orientation, it is reasonable to ask what would happen if we add *lead-by-orientation* to lead-by-attraction and lead-by-repulsion. We can then combine (a) repulsion-, attraction-, and orientation-style models with (b) various agent classes (human aware and human blind) and types (special, stakeholder, type blind, and type aware), and explore what happens.

Afforded Behaviors

The subsequent sections review prior work from our lab that explores various combinations of influence style and agent class. Although the sections are informative, it is important to note how many combinations haven't been studied. For example, how would type aware agents respond to the presence of a human aware agent that exerted a lead-by-orientation style.

Type Aware Agents When human blind agents are type aware (that is, they don't receive direct input from a human but do switch behaviors when in the presence of a leader), Couzin's phases tend to disappear. Type aware agents tend to line up behind a lead-by-attraction agent regardless of whether the agents started as a swarm, torus, or parallel group. The type aware agents near the leader start following the leader, pulling other agents along, and turning the entire group into a type of flock.

Type aware agents also tend to be fractured by a lead-by-repulsion agent, forming smaller groups. If there are enough initial agents in the collective, then some of these smaller groups can form a torus, swarm, or a parallel group as long as they are not near the predator.

Lead-by-attraction tends to work best when the task requires that agents move together and cluster around an object of interest, and lead-by-repulsion tends to work best when the task requires a lot of different agents to be distributed to different areas of the world (Goodrich et al. 2012). Naturally, since lead-by-repulsion tends to fracture the group, human workload is higher than with lead-by-attraction because the task shifts from one of influencing the entire group to managing a multiple smaller groups.

Type Blind Agents: Stakeholders versus Special Agents

The fact that type aware agents tend to lose structure in the presence of human-controlled agents is a problem. We can address this problem by exploring how type blind agents respond to the presence of human-controlled agents. We've compared how type blind agents respond to lead-by-attraction style influence using stakeholders or special agents (Goodrich et al. 2012).

Results indicate that, although it is possible to move a torus, swarm, or parallel group using either stakeholders or special agents, it requires fewer stakeholders to move the groups than special agents. Quoting from (Goodrich et al. 2012) and recognizing that the term "manager" in that paper is the same as "human aware" agent in this one, "For both manager styles, one and two managers had only a modest impact on the ultimate distribution of the phase, but as the percentage of managers grows their influence also grows. [Figure 3] shows how the error decreases much more rapidly for stakeholders than [special agents], and also that the apparent plateau of error is smaller for stakeholders than [special agents]."

Type Blind Agents: Switching Phases

The work reported so far either uses just one phase (flock-like structures, in the case of type-aware agents) or pulls a

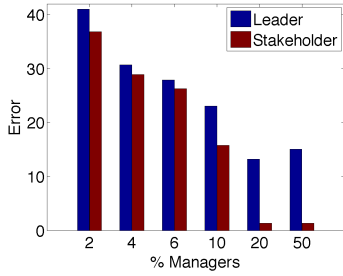


Figure 3: Error as a function of the percentage of number of managers. Error is Euclidean distance between the final centroid position and the food source. In the figure key, “Leader” means special agent. Figure taken from (Goodrich et al. 2012).

single structure from one point to another (in the case of type-blind agents).

Another line of research is enabling a human to shift between the two attractors with relatively little exertion of human influence. The research from (Kerman, Brown, and Goodrich 2012) indicates that it is possible to select agent parameters in such a way that a human can cause the collective to switch between structures. Preliminary work has indicated that, given a system tuned so that either parallel groups or tori are equally likely, it is easy to switch from a parallel group to a torus using lead-by-attraction. However, lead-by-attraction does not work as well for switching from a torus to a parallel group. This indicates that it is necessary to consider other forms of influence, and we have explored using lead-by-orientation. Preliminary results indicate that lead-by-orientation works very well for switching from a torus to a parallel group, as indicated in Figure 4.

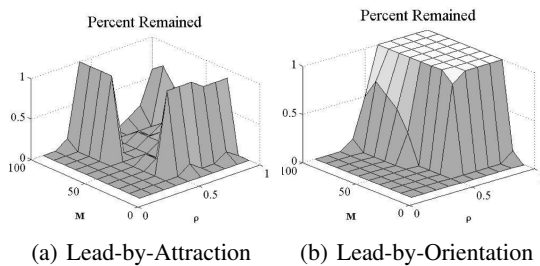


Figure 4: Probability of changing from a torus to a parallel group under different influence styles. On the axes, M indicates the number of stakeholders and ρ indicates how much weight the stakeholder places on human input.

Type Aware Agents: New Phases The previous discussion gives the impression that lead-by-repulsion turns all structures into either fragmented groups or into a flock. This does often happen, but recent preliminary results also indicate another possibility. When a special agent using lead-by-repulsion is placed in the center of a torus of type blind agents, the type blind agents spin around the special agent.

By moving the special agent, the torus can be “steered” by the agent inside of it.

This leads to an interesting possibility of allowing multiple special agents to use combinations of attraction and repulsion to cause the torus to take on various shapes. Preliminary results are shown in Figure 5.

Persistence and Span of Control

Consider the following example. When using special and type aware agents, it appears that it is possible for a human to exert sufficient control authority via only one special agent. The human can turn the team into a flock and then lead the flock to where he or she wants, or fragment the group and chase subgroups to where he or she wants. Thus, this approach allows a single locus of human control. The cost is that the human must persist in exerting this control over a period of time. For example, to chase a subgroup to a desired location, experiments with human subjects indicate that the humans are continually moving the predator in order to keep the subgroup together.

As illustrated by the previous example, control authority includes both span of control and persistence. In this paper, span of control is simply the number of human aware agents in the collective, and persistence is the amount of time that the human aware agents must exert authority. Thus, we can generalize the definition of interaction time from (Crandall et al. 2005; Olsen, Jr., Wood, and Turner 2004) to *interaction effort*, defined as the product of span of control and persistence. Interaction effort has a unit of *agent-seconds*, meaning that interaction effort is measured in the number of agents that must be influenced and for how long.

We hypothesize that each combination of influence style and distribution of agent types has a characteristic interaction effort. As evidence for this hypothesis, we note how both special agents and stakeholders can attract a structured team from one location to another, but fewer stakeholders are required. Persistence is frequently hidden in the results above; all figures presented above are shown after the swarm changed behaviors, which sometimes took considerable time.

Future work should seek to quantify the interaction effort for various team designs. If interaction effort can be quantified, even probabilistically, then there is reason to believe that this information could be used to provide performance guarantees for human-swarm teams (Goodrich and Mercer 2012).

Beyond Swarms

Although interesting, there is a major weakness in the work presented above: the agents cannot autonomously respond to environmental signals. The agents swarm, adopt shapes, can be lead, and can be coaxed into changing shapes, but the agents do not autonomously change their shapes or move toward particular locations in the environment. We refer to these agents as swarms precisely because their autonomy is designed to maintain a collective structure but not to do anything useful for the group.

Autonomy is enabled with $\epsilon(\cdot)$ from Equation (1) is non zero. Recall that we let u_t^{env} denote some spatially located

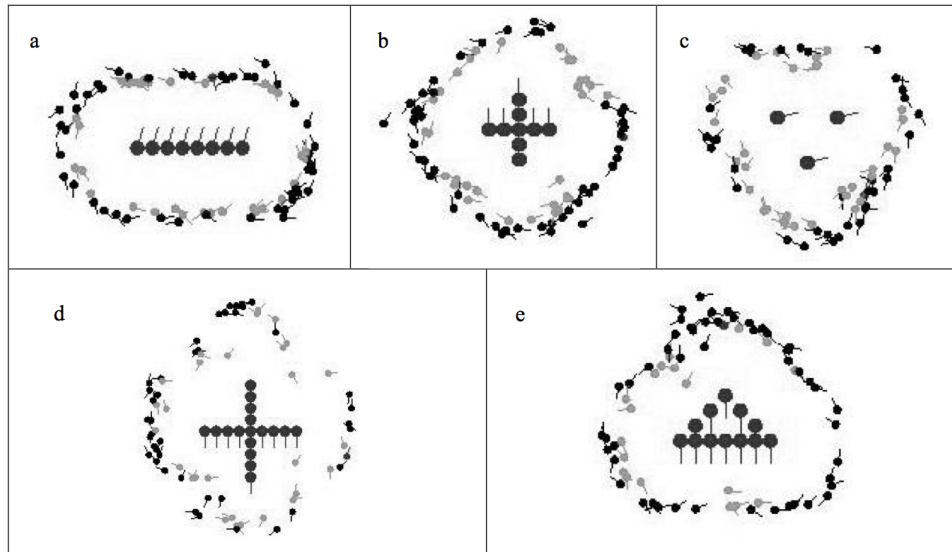


Figure 5: Special agents (the large circles in the center) can use a combination of attraction and repulsion to create various tori-like structures among type blind agents (small circles around the exterior). The type blind agents are using only attraction with and repulsion from each other, so the torus consists of agents circling both directions around the exterior.

signal in the environment, such as a food resource. If we extend x_t^i so that it denotes not only the kinetic state of the agent but also information about how agent i has reacted to u_t^{env} , then the set of ways in which a human can influence the collective grows. For example, in the problem of nest selection multiple agents could be exploring various locations as potential locations for a new nest. Agents leave the current nest to evaluate the potential site and then return to broadcast information about the quality of the site. The state of these agents include not only their current position, but also the location and quality of a potential nest site. Other agents could then commit themselves to exploring that site by observing this portion of the state. Notice how the parameter of the $e(\cdot)$ function in Equation (1) includes the state of all agents, allowing an agent to respond to a new environment signal by comparing it to the information from other agents.

We are just beginning work on these types of structures, but in reviewing the work in (Sumpter 2012) we propose to begin with the following. One way that animals coordinate activity in the wild when they aren't always within proximity of each other is by using a home, nest, or hive as a centralized place where they can communicate information. This creates a type of cohesion that has both spatial and temporal elements: spatial because communication occurs when agents are “with range” of each other which is most likely to happen when they are at the nest site, and temporal because different agents pick up different information at different times.

Based on Sumpter's work, we propose that the attractors for such nest-based collectives is a spatial distribution of agents, such as when ants or bees distribute themselves along various paths retrieving food from various sources. We don't know quite how to influence these collectives, but we propose that in addition to attraction, repulsion, and orienta-

tion, a key component of human influence will be the ability to inject information into x_t^i about how agent i has reacted to u_t^{env} .

Related Work

We have previously written literature reviews for human interaction with bio-inspired robot teams and refer the reader to those papers for discussions of that literature (Kerman, Brown, and Goodrich 2012; Goodrich and Mercer 2012; Goodrich et al. 2011a). In this section, we briefly review literature that distinguishes between swarms and more sophisticated types of animal collectives.

For Couzin's biomimetic and Spears' physicomimetic models, spatially cohesive groups are capable of expressing multiple, qualitatively different collective structures or phases. Wood and Ackland use a similar model as Couzin's, but they identify four different collective structures or phases that they label pack, swarm, dynamic, and no group. Wood and Ackland's swarms incorporate Couzin's swarms and tori, and the dynamic category incorporate Couzin's dynamic and highly parallel groups. Models from physics exhibit similar distinctions between relatively stationary swarms and more highly mobile flocks (Levine, Rappel, and Cohen 2000).

Conradt and Roper note that “Consensus decisions are made by *spatially cohesive groups* and usually concern movement direction, travel destination, and activity timing” (Conradt and Roper 2005, emphasis added). The spatial cohesiveness of the groups is emphasized because it is key. Flocks and swarms are spatially cohesive in the sense that they are structured as a spatially connected topology. By contrast, what we are calling colonies include how the so-called “eusocial insects” cited in (Conradt and Roper 2005) make decisions about nest choices. These choices include

the independent activities of foragers and scouts, which are then communicated to the remainder of the collective, triggering recruitment activities and quorum decision-making; we will model this using topological structures in a subsequent section.

Introducing heterogeneity adds another potentially useful dimension to swarm and colony-like behavior. For these problems, the payoff to each agent depends on simultaneously fulfilling different roles, suggesting the need to operate under a type of social contract (Skyrms 2004). Madden et al. recently created a model of pack activity that included different roles (Madden, Arkin, and McNulty 2010). Importantly, the different roles performed by members of the pack were dynamic and decentralized. Pack-level behaviors are perhaps the least studied in bio-inspired robot teams.

Conclusions

This paper has presented a synthesis of prior work, and identified a number of trends as well as areas needing more work. Human swarm interaction can be shaped by controlling how a human injects influence into a swarm (via human aware agents), the style of influence (attraction, repulsion, orientation, and possibly state information), and the interaction effort of the influence (persistence and span of control). Rich behaviors can be produced, but much work needs to be done, especially in extending results beyond simple swarms to colony- and pack-like behaviors.

References

- Conradt, L., and Roper, T. J. 2005. Consensus decision making in animals. *TRENDS In Ecology and Evolution* 20(8).
- Couzin, I. D.; Krause, J.; James, R.; Ruxton, G. D.; and Franks, H. R. 2002. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* 218(1).
- Crandall, J. W.; Goodrich, M. A.; Jr., D. R. O.; and Nielsen, C. W. 2005. Validating human-robot interaction schemes in multi-tasking environments. *IEEE Transactions on Systems, Man and Cybernetics — Part A: Systems and Humans* 35(4).
- Goodrich, M. A., and Mercer, E. G. 2012. Abstraction and persistence: Macro-level guarantees of collective bio-inspired teams under human supervision. In *Proceedings of Infotech@Aerospace*.
- Goodrich, M. A.; Sujit, P. B.; Pendleton, B.; and Pinto, J. 2011a. Toward human interaction with bio-inspired robot teams. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*.
- Goodrich, M. A.; Sujit, P.; Kerman, S.; Pendleton, B.; and Pinto, J. 2011b. Enabling human interaction with bio-inspired robot teams: Topologies, leaders, predators, and stakeholders. Technical Report BYU-HCMI Technical Report 2011-1, Brigham Young University.
- Goodrich, M. A.; Pendleton, B.; Kerman, S.; and Sujit, P. B. 2012. What types of interactions to bio-inspired robot swarms and flocks afford a human? In *Proceedings of Robotics Science and Systems*.
- Kerman, S.; Brown, D.; and Goodrich, M. A. 2012. Supporting human interaction with robust robot swarms. In *Proceedings of the International Symposium on Resilient Control Systems*.
- Levine, H.; Rappel, W.-J.; and Cohen, I. 2000. Self-organization in systems of self-propelled particles. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* 63(1).
- Madden, J.; Arkin, R. C.; and McNulty, D. 2010. Multi-robot system based on model of wolf hunting behavior to emulate wolf and elk interactions. In *Proceedings of the IEEE International Conference on Robotics and Biomimetics*.
- Olfati-Saber, R. 2006. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on Automatic Control* 51(3):401–420.
- Olsen, Jr., D. R.; Wood, S. B.; and Turner, J. 2004. Metrics for human driving of multiple robots. In *Proceedings of the 2004 IEEE Intl. Conf. on Robotics and Automation*, volume 3, 2315–2320.
- Reynolds, C. W. 1987. Flocks, herds and schools: "a" distributed behavioral model. In *SIGGRAPH '87 Proceedings of the 14th annual conference on Computer graphics and interactive techniques*.
- Sheridan, T. B., and Verplank, W. L. 1978. Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory.
- Skyrms, B. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.
- Spears, W.; Spears, D.; Heil, R.; and Kerr, W. 2005. An overview of physiocommimetics. In Sahin, E., and Spears, W., eds., *Swarm Robotics*, Lecture Notes in Computer Science. Berlin: Springer. 84–97.
- Sumpter, D. J. T. 2012. *Collective Animal Behavior*. Princeton, N.J.: Princeton University Press.