2013-09-04

Human-Swarm Interaction: Effects on Operator Workload, Scale, and Swarm Topology

Brian O. Pendleton
Brigham Young University - Provo

Human-Swarm Interaction: Effects on Operator Workload, Scale, and Swarm Topology

Brian Pendleton

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Michael Goodrich, Chair
Kent Seamons
Quinn Snell

Department of Computer Science

Brigham Young University

September 2013

ABSTRACT

Human-Swarm Interaction: Effects on Operator Workload, Scale, and Swarm Topology

Brian Pendleton
Department of Computer Science, BYU
Master of Science

Robots, including UAVs, have found increasing use in helping humans with dangerous and difficult tasks [61]. The number of robots in use is increasing and is likely to continue increasing in the future. As the number of robots increases, human operators will need to coordinate and control the actions of large teams of robots. While multi-robot supervisory control has been widely studied, it requires that an operator divide his or her attention between robots. Consequently, the use of multi-robot supervisory control is limited by the number of robots that a human or team of humans can reasonably control [57]. Swarm robotics – large numbers of low-cost robots displaying collective behaviors – offers an alternative approach by providing the operator with a small set of inputs and parameters that alter the behavior of a large number of autonomous or semi-autonomous robots. Researchers have asserted that this approach is more scalable and offers greater promise for managing huge numbers of robots [61].

The emerging field of Human-Swarm Interaction (HSI) deals with the effective management of swarms by human operators. In this thesis we offer foundational work on the effect of HSI (a) on the individual robots, (b) on the group as a whole, and (c) on the workload of the human operator. We (1) show that existing general swarm algorithms are feasible on existing robots and can display collective behaviors as shown in simulations in the literature, (2) analyze the effect of interaction style and neighborhood type on the swarm's topology, (3) demonstrate that operator workload stays stable as the size of the swarm increases, but (4) find that operator workload is influenced by the interaction style. We also present considerations for swarm deployment on real robots.

Keywords: HSI, Swarm, Robotics, HRI, Mental Workload, Human Factors, Topology, Neglect Time

ACKNOWLEDGMENTS

I sincerely thank my advisor, Dr. Michael Goodrich, for his knowledge, guidance, and instruction as I've completed my research and for his empathy and patience that made the journey much easier. Thanks are extended to the members of my committee, Dr. Kent Seamons and Dr. Quinn Snell. We also extend thanks to Dr. Snell in his role as graduate coordinator, to prior graduate coordinator Dr. Dan Ventura, and to department chair Dr. Parris Egbert, as well as to the many other faculty and staff of Brigham Young Unversity for their knowledge, instruction and time. Members of the HCMI and MAGICC labs, Brigham Young University provided fellowship, helpful ideas and suggestions. Dr. Dennis Eggett provided statistical expertise and analysis. I acknowledge the authors and maintainers of ROS, OGRE, CEGUI, Matlab, Ubuntu Linux, GCC, LaTeX and numerous other software packages that have helped me accomplish my research. I wish to thank my parents, Burton and Rosemary Pendleton for their editing help and support. Finally, I thank my wonderful wife, Rebecca, and children, Harold and Jace, for all the love, support and smiles that made this thesis possible.

This research was funded by the United States Office of Naval Research and Brigham Young University, Department of Computer Science. The results of this work represent the opinions of the author and not of those that funded the work.

Contents

List of Figures

List of Tables

Chapter 1

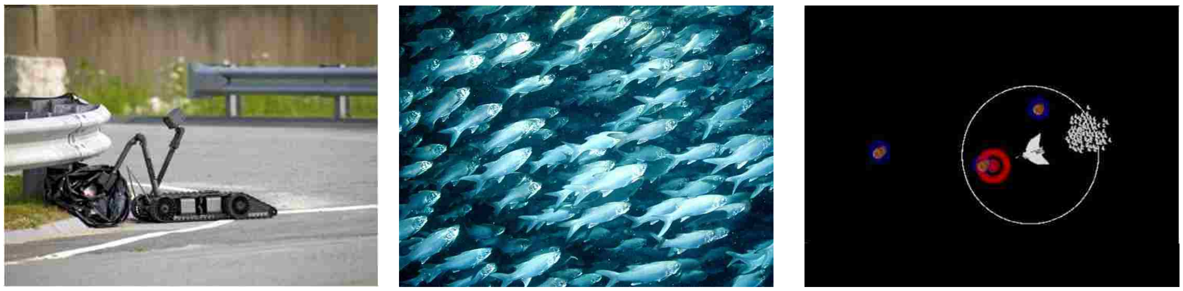
Introduction

In this thesis, we examine the effect of Human-Swarm Interaction (HSI) on a human-swarm system. In particular, we are interested in the mental workload of the human operator, the topology and stability of the swarm, and the effects on the individual robots. Section 1.1 provides information and context about the problem, why it's important, and how we propose to extend existing research. Section 1.2 presents the problem statement, which outlines the problem addressed by this thesis. Section 1.3 presents background, emphasizing definitions of key terms and presenting an overview of existing work. Section 1.4 presents the swarm models and modifications to those models used throughout this thesis. Section 1.5 contains a detailed literature review.

1.1 Motivation

Robots, including UAVs, have found increasing use in helping humans with dangerous and difficult tasks (see [61] and Figure 1.1a). The number of robots in use is increasing and is likely to continue increasing in the future. As the number of robots increases, human operators will need to coordinate and control the actions of large teams of robots. While multi-robot supervisory control has been widely studied ([53, 42, 24] for example), it requires that an operator divide his or her attention between robots, and consequently its use is limited by the number of robots that a human or team of humans can reasonably control [57].

In nature, large numbers of individuals frequently form large collectives or "swarms" [21]. Swarms allow many distinct individuals to function as a large cohesive unit even in the



(a) A bomb robot investigating a suspicious package

(b) A school of fish

(c) A graphical user interface for controlling a swarm

Figure 1.1: Robots, swarms, and human-swarm interaction

absence of leaders and centralized information (Figure 1.1b). Current work in the biology, physics and robotics communities has sought to understand the principles underlying these systems and encode them into robots. Current work in the emerging field of Human-Swarm Interaction (HSI) has shown that human operators can successfully interact with swarms of robots to perform useful tasks [9, 28, 39]. The body of HSI knowledge is expanding rapidly, but there are still many open questions.

Hypothetically, robot swarms that use simple decentralized rules are algorithmically scalable and can be managed by humans using a limited amount of shared information (see [9, 61] and Figure 1.1c). This is in contrast to traditional human supervisory control, where the operator must divide his or her attention between robots; consequently, as the number of robots managed by the operator increases, so does the operator's workload [57]. In a swarm system the operator is instead provided with high-level inputs that alter the behavior of the individual agents or the organization of the group.

While a swarm-based approach offers many benefits and useful properties, implementing swarm robotics presents a number of challenges. Swarm robotics research is a fairly young area, with many research problems that still need to be explored. This is especially true of human-swarm interaction, where research on single operator control is just emerging and multiple operator control remains largely unstudied[1].

[1]Future research on multi-operator control may draw from work by Conradt et al. regarding mixed goals in schools of fish [18]. Because swarm robots take advantage of interactions with their immediate neighbors,

One of the primary factors that limits the number of robots a human operator can control is the mental workload of that operator [57]. Mental workload is defined by Sanders and McCormick [62] as "a measurable quantity of the information processing demands placed on an individual by a task." We hypothesize that properly designed swarm systems can offer high-level inputs that don't require a human operator to switch attention between individual robots. We further hypothesize that this approach will prove more scalable in terms of operator workload; we are not aware of any studies that have attempted to evaluate this effect.

A variety of methods for interacting with swarms have been proposed. Many existing algorithms depend on specific robot hardware capabilities or depend on specific sensor configurations. In this thesis, we focus on simple high-level control methods that can be applied across a variety of robot and sensor configurations. We expand on published work by Goodrich et al. [33] on the impact of several simple high-level control methods (leader, predator and stakeholders) on the human-swarm system[2]. We examine the impact of these control methods on the topology of the swarm itself as well as on the workload of the human operator.

1.2 Problem Statement

In this thesis we offer foundational work on the effect of HSI (a) on the individual robots, (b) on the group as a whole, and (c) on the workload of the human operator. We (1) show that existing general swarm algorithms are feasible on existing robots and can display collective behaviors similar to behaviors reported in the literature, (2) analyze the effect of interaction style and neighborhood type on the swarm's topology, (3) demonstrate that operator workload stays stable as the size of the swarm increases, and (4) find that operator workload is influenced by the interaction style.

control can be exerted over the operator's local neighborhood, allowing robots to be smoothly split between one operator and the next without the need for complex protocols [32]

[2] The author's contribution to the publication is included as part of this thesis

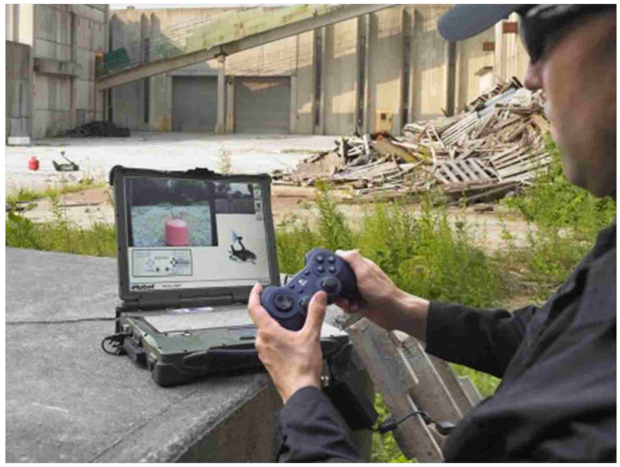


Figure 1.2: A tele-operated bomb disposal robot

1.3 Terminology

This section introduces key terms and ideas that are referenced throughout this document.

1.3.1 Human-Robot Interaction (HRI)

Goodrich [35] defines Human-Robot Interaction (HRI) as "a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans." HRI incorporates contributions from a variety of fields, including human factors, engineering, mathematics, computer science, robotics, cognitive psychology, and design with the goal of facilitating interaction between humans and robots, and designing effective human-robot systems [35]. Figure 1.2 shows a human remotely operating a bomb disposal robot using a graphical user interface and hand-held controller.

Human Supervisory Control

In human supervisory control of semi-autonomous robots, a human operator monitors the robot's execution of a task and periodically intervenes to operate the robot or reprogram it, for example giving it new objectives [64]. In a multi-robot system, human operators switch their attention between robots. The number of robots that a human operator can control with this paradigm is determined by the neglect tolerance of the robot – the amount of time that each robot can act autonomously before requiring human input – and the number of things that the operator can do at once [25, 57]. Operator workload is an important contributing factor to the number of tasks the operator can manage (see [57] and Section 1.3.5).

Information Foraging as a Canonical Problem

Information foraging involves one or more agents that move from location to location searching for, and "consuming," some type of resource(s). The location may be physical (e.g. rooms in an office building) or virtual (e.g. web pages). The "resource" is gathered at a finite rate and may or may not be depletable, depending on what it represents. The goal of the agents and, by extension, the human operator is to amass as much of the resource(s) as possible.

Information foraging represents an abstraction across many different types of problems. For example, in a UAV surveillance task, the UAV may move throughout the world to surveillance targets (locations), which come and go throughout the day. Time on target is the resource of interest, with the consumption rate proportional to the importance of the surveillance target. The goal of the UAV is to gather as much surveillance data as possible, adjusted by the importance of that data. The resource to be gathered can also represent a commitment on the part of the agent, such as the time taken to perform a task. Figure 1.3 shows simulated fish engaged in an information foraging task.

1.3.2 Swarm Robotics

What are Swarms?

In nature, large numbers of individuals frequently form large collectives or "swarms" [21]. Swarms allow many distinct individuals to function as a large cohesive unit even in the absence of leaders and centralized information. Examples of swarms include flocks of birds, schools of fish and ant colonies (Figure 1.4a). Swarms exhibit high-level emergent behaviors based on simple interaction rules between individuals. Research in swarm robotics seeks to encode collective behaviors found in nature into robots in order to allow large numbers of

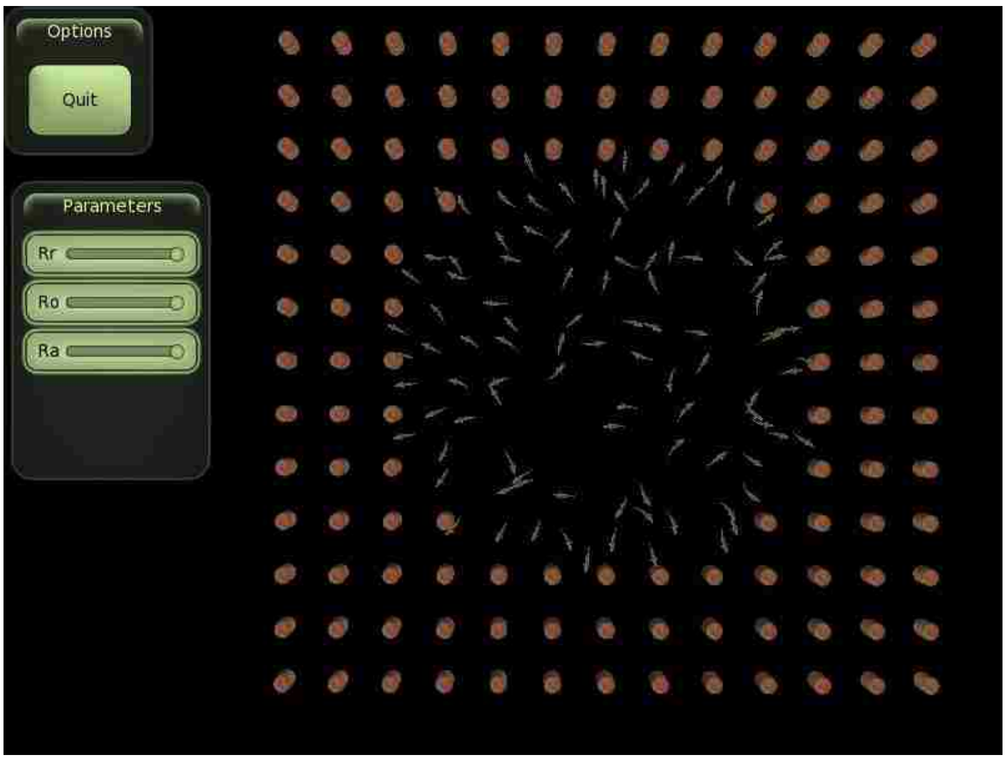


Figure 1.3: A simulated school of fish completing an information foraging task. The barrels represent information to be gathered, such as food or knowledge about the area in which they live.

(a) A swarm of ants form a bridge to cross a gap onto another leaf. Photo by Kasi Metcalf[3].

(b) A swarm of robots escapes its holding pen. The swarm can collectively accomplish goals that the individual robots could not[4].

Figure 1.4: Biological and robotic swarms

robots to collaboratively solve problems [61]. Researchers are interested in applying swarm principles to robots because of the useful properties they exhibit. Swarms are scalable, robust, do not require centralized leadership or information, and deal well with temporal-spatial problems and coordination between large numbers of individuals [61].

Swarm Properties

Swarms have several important and useful properties. First, swarms are scalable. Swarm algorithms use interactions among individuals in a local neighborhood, allowing them to scale from ten or twenty individuals to tens of thousands [61], as has been shown for starling flocks [7]. Second, swarms are robust. Because swarms contain no explicit hierarchy or designated leader, individuals can join or drop out from the group without substantial impact to large scale group behavior [12]. This is especially important in the challenging environments where individual robots may fail or be knocked offline. Third, swarms deal well with spatial problems using robust emergent behaviors [68]. In nature, swarms coordinate the movement

[3]Photo by Kasi Metcalf and licensed under the Creative Commons Attribution Non-commercial No-derivatives 2.0 Generic License. Original photo: `http://www.flickr.com/photos/kasimetcalfe/118471837/`

[4]Photo courtesy of the SYMBRIAN project: `http://symbrion.eu`

and behavior of large numbers of individuals over a physical area. Many of the rules distilled from natural systems have found their way into swarm robotic algorithms [61]. Last, because swarm robotics deals with interactions among individual agents, it also provides a framework to allow robots to react directly to operator state (such as movement and gaze) instead of exclusively requiring programmatic control inputs. One example is the GUARDIANS project [59] where swarm robots reacted to the movement of firefighters instead of requiring firefighters to enter programmatic inputs.

Application to Robotics

Swarm robotics seeks to take information learned from studying swarms in nature and encode that information into robots. The goal of this research area is to coordinate large numbers of relatively simple robots to accomplish larger tasks that none could accomplish alone (see figure 1.4b). Research in this area has focused on the construction of simple robots suitable for swarm algorithms, algorithms for controlling a swarm of robots, the engineering of emergent group behaviors such as formation control, and interaction and control of swarm systems.

1.3.3 Human-Swarm Interaction (HSI)

Human-swarm interaction (HSI) can be operationally defined as understanding, designing, controlling, and interacting with large-scale and decentralized autonomous or semi-autonomous systems. While the term "human-swarm interaction" was coined within the swarm robotics community [50], HSI traces its roots, not only to traditional HRI, but to the broad group of communities engaged in swarm research. The results and research are influenced by and contribute to biology, physics, robotics, and control theory, among others. Figure 1.5 shows a graphical user interface for human-swarm interaction.

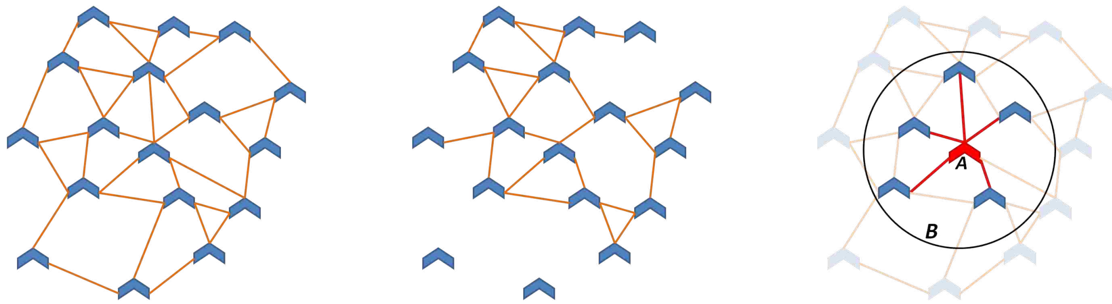
While traditional HRI focuses on individual robots or aggregates of individual robots, human-swarm interaction focuses on influencing or interacting with the group as a whole, or introducing environmental or group-level changes to affect the behavior of individual robots.

Figure 1.5: A user-controlled agent, marked with a white box, splits the swarm into two groups by repelling swarm members close to it.

Additionally, human-swarm interaction includes scenarios where the interaction rules of the group cannot be modified or are not completely understood (e.g. work by Marras and Porfiri on leading schools of fish using robots [48]).

1.3.4 Topologies and Neighborhoods

Swarm members typically interact with a small set of other swarm members and environmental factors that are close to them. This *local neighborhood* of interaction is responsible for many of the useful properties that swarms display [61]. This is especially true of flocking algorithms, which deal with the coordinated movement of the swarm. The overall structure of the interactions between swarm agents form a graph where the agents or environmental factors are nodes and the interactions are edges (Figure 1.6). This graph or *topology* describes the structure of the swarm and may be analyzed to determine qualities of the swarm or provide performance guarantees. The topology can be analyzed statically or over time. Graph theory provides useful tools for performing this analysis [51].



(a) Topological structure (b) Reduced interaction range (c) An agent (A) and its local neighborhood (B)

Figure 1.6: The topological structure of a group of simulated UAVs, where the ability to interact with surrounding UAVs is limited by distance. The UAVs are shown in blue, while the interactions are shown in orange.

The criteria for selecting a neighborhood, which influences the swarm topology, varies from algorithm to algorithm, but is a trait that all swarm algorithms share [61]. For example, Ballerini [7] found that flocks of sparrows react to the nine closest sparrows that they can see, while other models, such as Couzin's model of schooling fish [23] interact with all other agents within a certain distance. The precise topological structure used by animals is a current area of research in biology [20].

1.3.5 Human Factors

The goals of human factors research are to understand human capabilities, properties and limitations and apply them to people and systems interacting in a wide variety of human-machine systems. For example, many "ergonomic" devices sold in stores are designed to aid or facilitate interaction with a computer. Current human factors research traces its roots to aircraft design during the first and second world wars. Research then shifted from considering primarily the pilot to examining the aircraft and pilot as a human-machine system and focusing on issues such as the design of readouts and controls. Bainbridge [5] provides an instructive example of human factors considerations in a factory setting.

An understanding of human factors is critical for the design of interfaces and systems in the field of human-robot interaction [1]. If human-swarm teams are to be successfully deployed in the field, it is important to understand and quantify the effect of the swarm system and interface on human team members. In this research, we focus on the effect of interaction style and scale – the number of robots in the swarm – on the mental workload of the human operator.

Mental Workload

One important factor to consider in human-swarm systems is *mental workload*. Sanders [62] defines mental workload as "a measurable quantity of the information processing demands placed on an individual by a task." For example, simply reciting the alphabet would likely result in relatively low mental workload, whereas reading academic papers while juggling and counting to 100 would likely result in relatively high workload.

Mental workload measures are used in many different areas and are frequently used in the field of human-robot interaction. A high mental workload will increase the likelihood of errors and reduce overall performance [1]. In human-supervisory control, the mental workload of the human operator(s) is an important factor in determining how many robots can be controlled [57].

There are several common methods for measuring mental workload, including surveys, secondary tasks [16, 54], and behavioral entropy [52, 31]. Surveys, such as NASA-TLX, directly ask the operator a series of questions either during or after the task. Types of questions may include the perceived difficulty of the task, the areas where the operator felt the most pressure or the operator's confidence in his or her performance. A secondary task is an additional action, such as counting the occurrence of a word or doing simple math problems, that is not directly related to the operator's main task or objective. The difference in secondary task performance between experimental conditions can be used to infer the workload of the operator. Behavioral entropy is an objective measure of the difference in operator behavior between two experimental conditions; see Boer [11] for a discussion of behavioral entropy and Goodrich et al. [31] for an application of behavioral entropy to HRI research.

1.4 Modeling

In this section we present the swarm model used throughout this thesis, the criteria for selecting it and modifications to the model to add control inputs.

1.4.1 Swarm Models

A variety of different swarm models and algorithms have been proposed. Models have come from biology, physics, control theory, robotics and several other academic communities. Swarm models provide system dynamics and algorithms for generating coordinated collective behavior among agents. Some models are general or theoretical, while others model a particular type of animal, robot, or abstract agent.

While a variety of different types of algorithms exist, in this thesis we will focus on flocking algorithms, which deal with the coordinated movement of swarm members. Movement of the group is common to many different problem types and is relatively easy to abstract across types of robots.

1.4.2 Model Selection Criteria

The criteria for selecting a model were as follows:

- **Existing Algorithm**: This allows us to focus our efforts on providing additional HSI contributions rather than adding another model to the extensive number found in the literature.
- **Simple**: Swarm algorithms allow relatively incapable robots to accomplish interesting tasks collectively. Thus, the algorithm selected should apply to relatively incapable robots, broadening the applicability of our work.

Figure 1.7: An example of a collective behavior displayed by a group of individuals. A group of fish swim together in a torus formation. Work by Couzin et al. [23] has shown that this type of collective behavior can arise from simple interaction rules between fish[5].

- **Feasible**: The algorithm should be applicable and possible to implement on both ground robots and UAVs using available sensors.
- **General**: The model should not encode any task-specific information and should be flexible enough to apply to multiple problem types, broadening the work's applicability.
- **Decentralized**: Several important swarm properties, such as scalability and robustness, depend on the local nature of swarm algorithms. As such, the algorithm should not make use of centralized information, such as the location of the swarm's centroid, shared between agents.
- **Interesting collective behaviors**: One of interesting features of swarms is the ability to generate interesting collective behaviors from relatively simple rules.

1.4.3 Couzin's Model

We selected a model published by Couzin et al. [23], which models schooling behavior in fish. This model was constructed using empirical observations of schooling fish and shows how common school behaviors can emerge from local interactions (see Figure 1.7). It is capable of several different types of collective behaviors, but we focus on only one: flocking/schooling.

Couzin's model uses a simple switching controller based on distance for control, and fixed speed with constant turning rate for dynamics (similar to Dubins airplane [17]). Additionally, Couzin's model is decentralized and is not tied to any specific sensor or sensing capability. Based on this, we chose this model as a good fit for our requirements. A full overview of Couzin's model can be found in [23], while [18] provides extensions to the original model, which we use to provide control inputs.

While Couzin's original model is in three dimensions, in this thesis we will confine the agents to a two-dimensional coordinate space, with the assumption that agents plan in 2.5 dimensions (x, y and height above ground) and change altitude to avoid collisions. We made this decision based on early experiments with control of the agents and to avoid issues with perspective and control on a two-dimensional computer screen. Additionally, the two-dimensional coordinate space more accurately models the execution of area coverage tasks by unmanned areal vehicles (UAVs) where the UAVs travel at a relatively constant height above ground and spread out to maximize sensor coverage.

1.4.4 Couzin's Model – Mathematical Description

Couzin's model considers a set of agents, $i = 1, 2, \ldots, N$, which move around a simulated world in discrete timesteps τ with the following dynamics:

$$\begin{aligned} x_i[t+\tau] &= x_i[t] + s\tau\cos\phi_i \\ y_i[t+\tau] &= y_i[t] + s\tau\sin\phi_i \\ \phi_i[t+\tau] &= \phi_i[t] + u_i^\phi \end{aligned} \tag{1.1}$$

where $[x_i, y_i]^T \in \mathcal{R}^2$ is the ith agent's position, $\phi_i \in [-\pi, \pi]$ is the agent's heading relative to the global x axis (east), s is the forward speed, and θ is the maximum turning rate. We

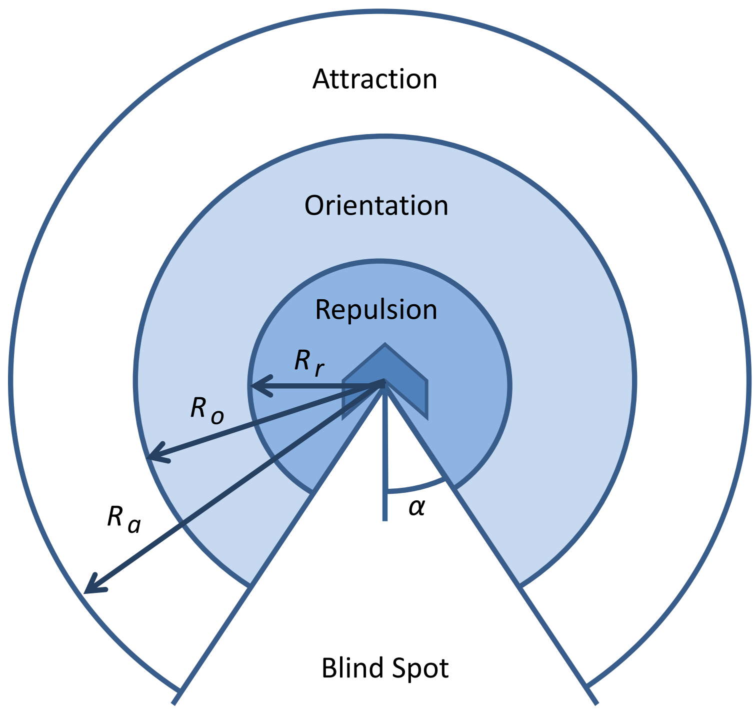


Figure 1.8: Couzin's model of schooling fish. Each simulated fish has three concentric zones of interaction and interacts with all other fish within these zones, provided they are not in the blind spot behind the fish. The fish is repelled by any other fish in the repulsion zone, attracted to those in the attraction zone, and tries to align itself with those in the orientation zone. The fish move forward with a constant velocity s and turn towards their desired heading at a turning rate θ.

define a position vector c_i, and velocity vector v_i for simplicity.

$$\begin{aligned} v_i = &\quad [\cos(\phi_i), \sin(\phi_i)]^T \\ c_i = &\quad [x_i, y_i]^T. \end{aligned} \tag{1.2}$$

As illustrated in Figure 1.8, the number of other agents in the three concentric zones of interaction is given by:

$$\begin{aligned} n_i^r &= \{j : 0 < \|c_i - c_j\| \le R_r, |\psi_{ij}| < 180^\circ - \alpha\} \\ n_i^o &= \{j : R_r < \|c_i - c_j\| \le R_o, |\psi_{ij}| < 180^\circ - \alpha\} \\ n_i^a &= \{j : R_o < \|c_i - c_j\| \le R_a, |\psi_{ij}| < 180^\circ - \alpha\} \end{aligned} \tag{1.3}$$

where ψ_{ij} is the angle between v_i and $c_j - c_i$. The model is a vector summation model where a desired heading is computed for each zone:

$$u_i^r = -\sum_{n_i^r} \frac{(c_j - c_i)}{|c_j - c_i|} \tag{1.4}$$

$$u_i^o = \sum_{n_i^o} \frac{v_j}{|v_j|} \tag{1.5}$$

$$u_i^a = \sum_{n_i^a} \frac{(c_j - c_i)}{|c_j - c_i|}. \tag{1.6}$$

Each agent i uses a switching controller to combine the desired heading from each zone and determine its overall desired heading d_i.

$$d_i = \begin{cases} u_i^r & \text{if } n_i^r > 0 \\ \frac{u_i^o + u_i^a}{|u_i^o| + |u_i^a|} & \text{if } |u_i^o| > 0 \text{ or } n_i^a > 0, \text{ and } n_i^r = 0 \\ v_i & \text{otherwise} \end{cases} \tag{1.7}$$

It then calculates its angle relative to the desired heading $\phi_i^d \in [-\pi, \pi]$.

$$\phi_i^d = \operatorname{atan2}(d_i^y, d_i^x) - \theta_i \tag{1.8}$$

and turns toward the heading with maximum turning rate θ.

$$u_i^\phi = \begin{cases} \phi_i^d & \text{if } |\phi_i^d| \leq \theta\tau \\ \theta\tau & \text{if } \phi_i^d > \theta\tau \\ -\theta\tau & \text{if } \phi_i^d < -\theta\tau \end{cases} \tag{1.9}$$

Noise is added by deviating the heading of the agent using a spherically wrapped Gaussian distribution with variance σ_g^2.

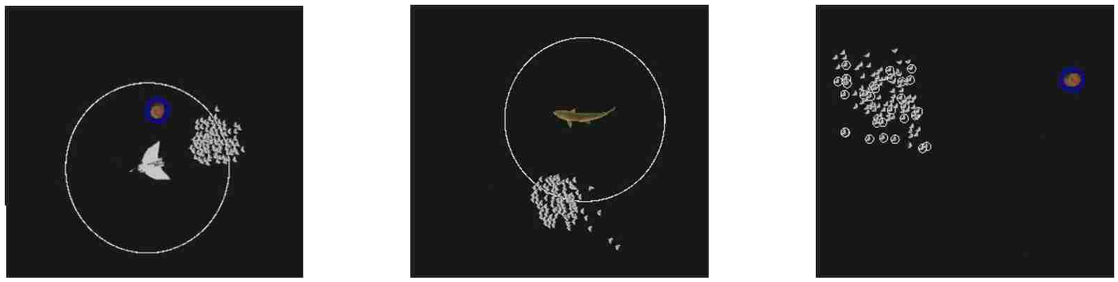
Parameter	Description	Units
s	Forward velocity	Units per second
θ	Turning rate	Degrees per second
α	Blind spot angle	Degrees
R_r	Zone of repulsion radius	Units
R_o	Zone of orientation radius	Units
R_a	Zone of attraction radius	Units
σ_g^2	Heading noise variance	None

Table 1.1: Summary of the parameters used by Couzin's model.

1.4.5 Control Methods

Many existing approaches for swarm control are tied to specific algorithms or robot capabilities, while others, such as potential fields and machine learning, are general, but require that the desired behavior be known a priori. In this research, we will instead focus on several simple control primitives that can be applied to a wide range of swarm algorithms. This approach provides foundational results that can be built on by future work and broadens the applicability of our results.

The three control primitives we investigate are: lead by attraction, lead by repulsion, and group influence. We refer to them as "Leader", "Predator", and "Stakeholders" respectively.



(a) Leader (b) Predator (c) Stakeholders

Figure 1.9: shows the three control methods used in this thesis, as shown in the graphical user interface. Leaders (a) are operator-controlled agents that cause other swarm members to turn towards them. Predators (b) are similar to leaders, but cause other swarm members to turn away from them. Stakeholders (c) are members of the swarm that are influenced by the human operator. They interact normally with other swarm members.

Leader

A leader is a physical or virtual agent that attracts all other swarm members within a radius of influence towards it. Using leaders, human operators only need to manage one swarm member for any swarm size. However, this approach requires swarm members to recognize leaders as distinct from other swarm members and may introduce a single point of failure without careful implementation.

Predator

Predators are similar to leaders, but repel swarm members within their radius of influence rather than attracting them. Leaders and predators have similar advantages and disadvantages. Additionally, work by Goodrich et al. [32], including work by the author, has shown that predators are able to split the swarm into subgroups more easily, but do not sustain influence as well as leaders (see Section 3.4).

Stakeholders

Stakeholders are regular members of the swarm that are directly influenced by the human operator. Stakeholders are not recognized as different by other members of the group, meaning that stakeholders can be anonymous and need not be the same agents at each time step. Additionally, work by Kerman, Brown, and Goodrich [44] shows that stakeholders can be used to to change the collective behavior of the group. However, stakeholders must comprise a substantial fraction of the group (generally 20-40% [44]).

1.4.6 Adding Human Influence to Couzin's Model

Couzin's model is a model of schooling fish and thus does not provide control inputs from the human operator. However, follow-on work by Conradt et al. [18] adds an additional term that directs the agents toward a goal location by weighting the existing model with a second vector that points toward the goal. We use this method for adding in the three methods of human influence. Additionally, we add a bounding box term using the same method to keep the agents within the view of the operator.

Adding Control Inputs

We augment Couzin's model with control inputs by weighting the original desired heading d_i from Section 1.4.4 with control vectors which represent (a) influence from the leader or predator u_i^p, or (b) a goal location only visible to stakeholders u_i^g, as follows:

$$\overline{d_i} = d_i + w_g u_i^g + w_p u_i^p \tag{1.10}$$

While these modifications allow the swarm to be influence by multiple control methods simultaneously ($w_g > 0$ and $w_p > 0$), we use only one control method at any time ($w_g > 0$ or $w_p > 0$, but not both). We also note that the agent's repulsion radius overrides external influence (e.g. $u_i^p = u_i^g = 0$ if $n_i^r > 0$). Table 1.2 contains a description of the control parameters added to Couzin's model.

Parameter	Description	Units
w_g	Stakeholders control gain	None
w_p	Leader/predator control gain	None

Table 1.2: Summary of the control parameters added to Couzin's model.

Leader and Predator

The influence from the leader or predator u_i^p consists of a bounded circular area, centered on the leader or predator. Any agent within the area of influence is either attracted toward the center (leader) or repelled toward the edge (predator), as follows:

$$n_i^p = \{j : \|c_i - c_p\| \leq R_p\} \tag{1.11}$$

$$u_i^p = -\sum_{n_i^p} \frac{(c_p - c_i)}{|c_p - c_i|} \tag{1.12}$$

where c_p is the position of the leader or predator. In the case of a leader, the sign of u_i^p is reversed. Note that the special case $w_p = \infty$ causes a switching behavior where agents ignore all other agents when within zone of influence of the leader or predator.

Stakeholders

Stakeholders are regular members of the swarm that are influenced by an external control signal u_i^g. In this thesis, we use the location of the participant's mouse (x_m, y_m) as a goal location which draws the stakeholders towards it, implemented as follows:

$$m = [x_m, y_m]^T \in \mathcal{R}^2 \tag{1.13}$$

$$u_i^g = \begin{cases} \frac{m - c_i(t)}{|m - c_i(t)|} & \text{stakeholders} \\ [0,\, 0]^T & \text{others} \end{cases} \tag{1.14}$$

Bounding

To facilitate our user studies, we introduce a bounding term b_i which is used to keep the agents within a playing field bounded by the participant's field of view. If an agent i is outside of the playing field, its desired heading is weighted with a bounding vector b_i, perpendicular to the playing field boundary the agent has crossed.

$$d_i^b = \frac{b_i + \overline{d_i}}{|b_i| + |\overline{d_i}|} \tag{1.15}$$

The angular gain is then calculated as in Equation 1.7, using d_i^b, the bounded control vector, in place of d_i.

1.5 Literature Review

1.5.1 Overview

The current understanding of swarms, swarm robotics and human-swarm interaction has come from a variety of different research communities including biology, mathematics, control theory, human-robot interaction, and human factors. Figure 1.10 shows the relationship between fields contributing to swarm robotics.

1.5.2 Biology

Work in the biology community has provided insight into models and mechanisms for self-organizing systems. Sumpter provides an excellent overview of some of the mechanisms that give rise to collective behaviors in biological systems [67]. Work by Couzin and associates has provided a survey of leadership and movement in animal groups [21], a demonstration of group behavior change as a result of individual changes [23], and a model of swarm control without explicit leadership [22]. These biological models have found application in robotics and inspired additional algorithms. Further work by Conradt and Couzin has explored conflicting goals within groups [18]. Ballerini examined flock structure in starlings and found

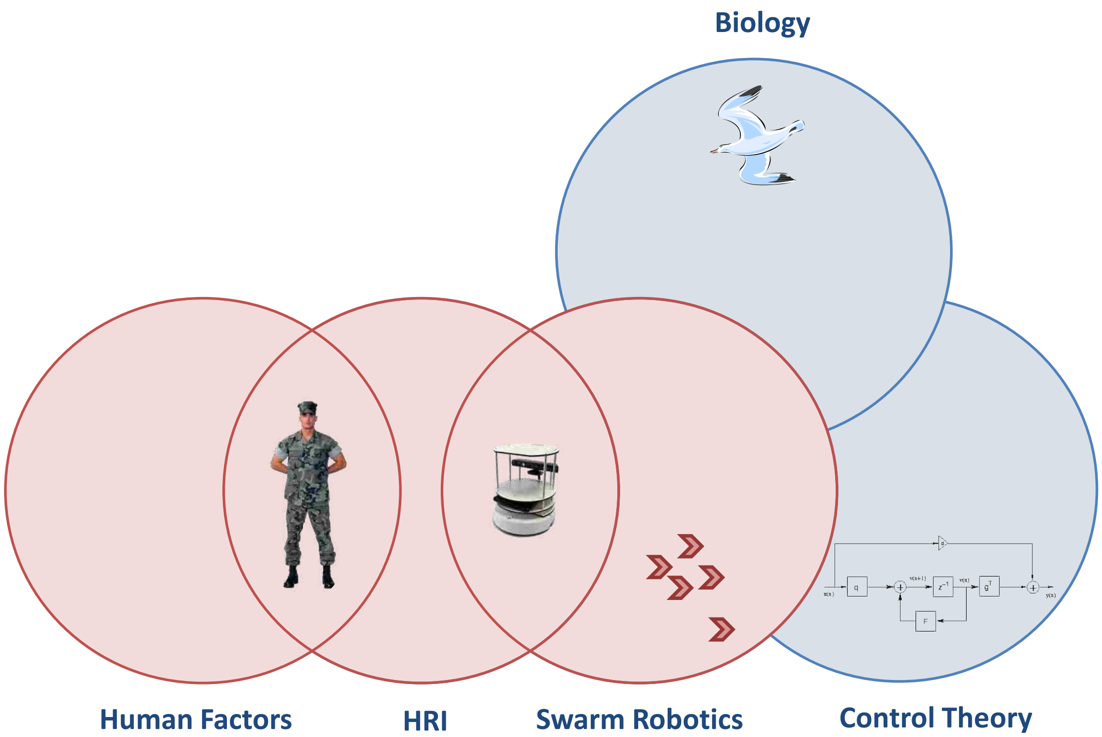


Figure 1.10: shows the relationship between the fields and literature contributing to this thesis. We draw primarily from HRI, human factors and swarm robotics and focus our contributions in human-swarm interaction.

evidence of anisotropic interactions and an interaction neighborhood based on topological rather than metric distances [7]. Our work draws heavily on the biology community for our model and many of the ideas found in this thesis. We use Couzin's model of schooling fish, and base our extensions to the model on the work of Conradt et al. We also incorporate ideas and results from Sumpter and Ballerini into the design of our experiments.

1.5.3 Mathematics and Control Theory

Researchers in the areas of mathematics and control theory have contributed mathematical models and control theoretic guarantees. The literature is extensive, so we only present a sample of the literature that directly relates to our work. One example is provided by Olfati-Saber's work on mathematical models and proofs of performance [55]. Barnes et al. used mathematical functions for swarm formation control [8]. Haas uses a similar algorithm to maintain a perimeter in a convoy escort task [38]. Work by Gray has demonstrated robot swarming without explicit communication [37]. Kira and Potter used a potential field model for controlling a robot swarm and evolved field weightings using a genetic algorithm [46]. We also recommend two books, *Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications* [60], and *Graph Theoretic Methods in Multiagent Networks* [51], which provide a more extensive overview of the literature. We draw on work in this area in the design of our control methods and development of considerations for deploying generic algorithms on real robots.

1.5.4 Human Factors and Human-Robot Interaction

As with many robotics applications, one of the goals of swarm robotics is to help people perform difficult and dangerous tasks. Research in human factors and human-robot interaction have provided understanding and evaluation methods for interfaces and human performance in human-robot teams. Goodrich and Schultz provide an excellent survey of human-robot interaction [35]. Adams provides important considerations for human-robot interfaces [1].

Bainbridge discusses challenges for humans working with automated systems [5]. Steinfeld et al. present common metrics for human interaction [65]. Common human factors metrics are discussed by Endsley (situation awareness) [27], Donald (vigilance) [26], and Sanders (mental workload) [62]. Amrein provides soldier characteristics and human factors considerations [3]. Bainbridge shows that humans react differently to robots that are physically present, instead of projected [6]. Because our research examines the impact of HSI on both the swarm and the human operator, we make extensive use of many human factors and HRI concepts in our research, particularly in Chapter 4. Our work incorporates concepts such as mental workload, neglect time, secondary tasks, and HRI interface design, to name a few. Much of our work can be viewed as an evaluation of HRI concepts in the context of human-swarm systems.

1.5.5 Human-Swarm Interaction

Existing research has also explored human-swarm interaction and methods for providing input to a swarm. Kira and Potter discuss how a robot swarm can be controlled using external or internal controls [46]. Alboul presents algorithms to allow firefighters to lead a formation of swarm robots as part of the GUARDIANS project [2]. Bashyal and Venayagamoorthy provided users with a virtual avatar and simple behavior primitives for influencing swarm behavior [9]. Work by Couzin shows how agents can lead a group even though they are not explicitly designated as leaders [22]. Prior research by Goodrich and associates at Brigham Young University, including work by the author of this thesis and included herein, has examined the use of leader and predator agents to control a swarm [34] and the use of agents within the swarm to change the formation (phase) of the group [36]. Our research makes use of this prior work to implement control strategies and demonstrate their use with the swarm.

Methods for designing swarm interfaces have also been explored in the literature. McLurkin et al. found that encoding swarm state information as lights and sounds allowed programmers to understand swarm behavior and find algorithmic errors more quickly [50]. Gancet et al. presented a swarm interface for use by firefighters as part of the GUARDIANS project [30]. Hass designed and evaluated an interface for solider-swarm interaction in a convoy task and found multi-modal interaction to be effective [39]. Bayshal and Venayagamoorthy designed an interface that allowed a user to select behavioral primitives to direct the swarm [9]. We incorporate many of the ideas presented by these authors into the design of our user study interface (Chapter 4).

Most existing swarm research uses simple homogeneous robot models and behaviors in simplified environments. Additional research is needed to explore more complex behaviors using realistic robot models. Existing research has provided useful spatial behaviors for swarms, but additional work will be needed in the future to provide for more complex behaviors in a spatial context. By far the greatest challenge is swarm engineering – the design of local interaction rules that reliably result in the desired high-level behavior [71] – due to the difficulty of predicting group-level behaviors from individual interaction rules and stochastic local behavior. However, some successful techniques are beginning to emerge [71]. In the robotics community, Kira et al. [46] were able to use machine learning algorithms in order to evolve the desired behaviors. Work in the math and control theory communities has provided control-theoretic guarantees for some swarm algorithms [55, 72]. We extend this prior work by demonstrating how high-level generic algorithms can be successfully encoded into robots (Chapter 2) and by evaluating the affordances and tradeoffs provided by three control methods (Chapter 5).

1.5.6 Summary

Human-Swarm Interaction is an newly emerging research field. As such, there is a lot of foundational work left to be done. Existing research has primarily focused on developing new algorithms and control methods, and on demonstrating that human control is feasible. While some studies have incorporated a human factors component, we are not aware of any studies that have attempted to compare the workload of the human operator across scale and control method or to analyze the impact of HSI on the swarm itself. The work described in this thesis will make a foundational contribution in this area.

Chapter 2

Individual Swarm Robots: Effects and Considerations

2.1 Introduction

In this chapter we show that Couzin's model (Sections 1.4.4 and 1.4.6) is suitable for use on physical robots, and that physical robots are capable of displaying the collective behaviors produced by Couzin's model in simulation. Additionally, we also examine considerations for deploying existing algorithms on real robots and some of the assumptions that are frequently encountered in simulated models. The work presented in this chapter serves as a concrete basis and justification for our later work and user studies, presented in Chapter 3, 4, and 5. It demonstrates that the experiments outlined in later chapters are feasible, and provides justification for our simulations and later assumptions and modeling of real robots. It also shows how a general swarm algorithm can be applied to physical robots successfully.

Successful implementation of existing swarm algorithms on general purpose robots requires consideration of the implementation of the algorithm and any assumptions that it makes. While some work as been done on generic libraries of algorithms for swarming robots (e.g. McLurkin [49]), many swarm algorithms that have been deployed on physical robots are developed in tandem with the robots themselves and take into consideration specifics of the robots' capabilities, sensors, and locomotion. Other more theoretical work has provided a large number of generic algorithms that have been shown to display swarm behaviors in simulation or proven to have certain properties, but have not been demonstrated on physical robots.

Couzin's model of schooling fish, which we use in this thesis, provides simulation results that demonstrate how common structures demonstrated by schools of fish can emerge from a simple, parameterized model. It was not demonstrated physically, due to the obvious difficulty in reprogramming fish to follow the model[1]. Rather than reprogramming fish, we demonstrated the feasibility of the model on two TurtleBots from Willow Garage (Figure 2.1a). We then used parameters gathered from the physical robots, along with the Stage robot simulator, to show that the collective behaviors described by the model are possible on a larger number of robots.

This chapter is organized based on the order of the experiments we conducted. Section 2.2 provides considerations and assumptions that were addressed to implement the algorithm on physical robots. Section 2.3 presents our initial experiments on the physical robots, demonstrating that swarming, proximate interaction, and Couzin's model are feasible on the TurtleBots. Section 2.4 explains how localization error and other parameters were gathered from the robots. Section 2.5 uses the parameters obtained from the physical robots and the Stage robot simulator to show that physical robots can demonstrate the collective behaviors shown by Couzin's model in simulation.

2.2 Considerations

We identified three main considerations for implementing Couzin's model on the TurtleBots. First, we needed to define frames of reference and establish how the TurtleBots would determine the relative location of other robots and a proximate human operator. Second, we needed to determine what communication was needed and verify that it could be accomplished with the robot's hardware. Third, we needed to evaluate the capabilities of the robot hardware to ensure it was sufficient to run the algorithm. A careful consideration of these issues allowed us to successfully implement Couzin's model on the robots and should aid those porting other swarm models to robot hardware.

[1]Current work may soon overcome this limitation, see Latif et al. [47].

2.2.1 Frames of Reference

Robots in a swarm react to information that they obtain about other robots in their local neighborhood. In the case of spatial information, such as position, orientation, and velocity, each robot must be able to resolve other robots relative to itself. In other words, each robot must be able to calculate a relative transformation from its own frame of reference to the neighboring robot's frame of reference. This can be accomplished in two ways: (1) the robot is able to directly resolve the relative position of other robots using its own sensors, or (2) the location of both robots is known relative to a third frame of reference, which the robots share. This third frame of reference could be another robot, a landmark they are both aware of, or a global coordinate frame, such as GPS. In case (1), the robot can use the information directly, while in case (2), the robot must receive information about the location of the other agent relative to the common reference. We call these two cases (1) *sensor-based resolution* and (2) *localize and communicate*.

In simulations, robots are typically positioned relative to a global "truth" coordinate frame, and know or calculate the position of themselves and other robots relative to the global frame, with some degree of noise. In real-world systems, GPS often provides a similar global frame of reference that allows autonomous systems to determine their relative position to each other or to landmarks. However, many important problem areas in robotics research take place indoors, where GPS is not typically available. Additionally, even when available, GPS may not always be usable due to jamming, reliability issues, or limited accuracy. In these situations, determining the relative transform between two agents can be difficult. If the robot does not posses sufficient sensor and processing capabilities to reliably track other robots, the two robots must agree on a shared frame and communicate their position relative to this frame. Recent work in online multi-robot SLAM provides one solution for this problem (see [63, 15, 45] for example). However, a full implementation of these algorithms is beyond the scope of this thesis.

2.2.2 Communication

Swarms use decentralized interactions to scale to a variable number of agents and avoid single points of failure. Interactions communicate information between swarm members and may involve direction communication between agents or indirect communication via the environment. In the case of HSI, communication also occurs between the swarm and one or more human operators. The communication system must be designed so that the locality of interaction is maintained (avoiding single points of failure) and so that the addition of more agents does not introduce failures or overwhelm the available bandwidth.

Communication between agents can be implicit or explicit, and can take place directly or through the environment. In the case of implicit communication, such as swarm-mounted lights or camera systems, swarm designers must ensure the communication is reliable and consider issues such as occlusion and environmental noise. In the case of explicit communication, such as radio beacons or mesh networks, swarm designers must carefully consider the topology and amount of information exchanged so that the communication scales to large swarm sizes and does not overwhelm the available bandwidth.

Global information, such as information passed from the swarm to the human operator, must be evaluated in similar fashion. Information about the location and state of swarm agents is of special interest, as many of the swarm user studies to date assume full observability of the swarm. Full observability requires increasing bandwidth as the size of the swarm increases, and can potentially overwhelm available bandwidth for larger swarm sizes. Future work should evaluate these bandwidth issues and the effect of partial observability on the human-swarm system.

2.2.3 Robot Capabilities

The capabilities, sensors, locomotion, and processing power of the robot platform determine which algorithms it can run. For example, algorithms that require robots to hold position are not suitable for UAVs, which must maintain forward velocity to stay airborne. It is also important to consider the information required by the algorithm to ensure that the robots can obtain it, either from their own sensors or another source, and can do so with sufficient accuracy and reliability. For example, algorithms that require location or heading information about adjacent agents require that the robots are capable of tracking other robots with their own sensors, or communicating their location with respect to a shared reference frame. These issues are especially important to swarm robotics, where robot platforms are specifically designed to low-cost and, consequently, have limited capabilities.

2.3 Proximate Interaction and Swarming

In this section, we discuss the experiments we performed to evaluate the feasibility of Couzin's swarm model and human-swarm interaction on two Willow Garage Turtlebots. Our goal was to evaluate (1) the sensing and communication abilities of the robots, (2) the interaction with both proximate and remote human operators, and (3) the implementation of Couzin's model on the robots. We conducted experiments to evaluate each one of these areas and present our results as follows: Section 2.3.1 discusses the Turtlebot hardware and methods used for localization and communication. Section 2.3.2 presents our experiments on proximate and remote HSI using the TurtleBots. Section 2.3.3 describes our tests of Couzin's model on the TurtleBots.

2.3.1 Robot Platform

We used two TurtleBots (Figure 2.1a) running the Robot Operating System (ROS) software, both created by Willow Garage. The TurtleBot hardware consists of an Asus Eee PC 1215n laptop, an iRobot iCreate base, a Microsoft Kinect sensor, a power module, a single-axis gyro, and mounting hardware. The robot's physical specifications are given in Table 2.1. We used the standard ROS navigation stack, which uses gmapping for SLAM, AMCL for localization against a known map, and an extended Kalman filter for sensor fusion. We used

Parameter	Description	Value
s	Forward velocity	$0.5 m/s$
θ	Turning rate	$90°/s$
α	Blind spot angle[3]	135.0°
R_{max}	Maximum sensor range[4]	4-5m

Table 2.1: The physical parameters of the Willow Garage TurtleBots used in the experiment.

the standard parameter values supplied with the TurtleBot, with the exception of gyro and encoder calibration, which were conducted as per the supplied instructions.

The experiment was conducted on the second floor of the TMCB at Brigham Young University. The TurtleBots communicated via a wireless network, which provided realistic real-world communications, including latency and communication dropouts[2]. Due to the limited field of view of the Kinect sensor (±45°), the Turtlebots used a self-localize and communicate approach to determine relative distance between them. Both robots localized against a known map of the 2nd floor and relayed their relative position. Online multi-robot SLAM with shared reference frames has been demonstrated [63, 15, 45], but an implementation of this work is beyond the scope of this thesis. We simply acknowledge this work and assume that a given robot and each of its neighbors can periodically arrive at a shared frame of reference relative to themselves with some degree of error. This frame of reference may differ between pairs of robots and need only be updated periodically. Localization against a SLAM-created map provides real localization error, similar to that which would be observed during multi-robot SLAM.

2.3.2 Remote and Proximate Human-Swarm Interaction

We conducted an experiment to validate that the TurtleBots could be used in proximate and remote HSI scenarios. The experiment utilized both TurtleBots in a simulated force

[2]Mesh network approaches have also been demonstrated (e.g. Correll et al. [19]).

[3]The blindspot is effectively eliminated due to a forward-facing sensor, constant forward motion, and a persistent occupancy grid. The field of view of the Kinect sensor is sufficient to localize accurately and identify obstacles in the direction of travel.

[4]Obstacles at a distance of more than 4m are visible, but with greatly reduced distance accuracy.

(a) TurtleBot (b) Wiimote

Figure 2.1: The hardware used in our evaluation of remote and proximate HSI using the TurtleBots. We used two Willow Garage Turtlebots (a) in a simulated force protection scenario. The human operator was equipped with a Wiimote (b) that could be used to give commands to the robots.

protection scenario involving a human operator. The two TurtleBots were responsible for escorting the operator as he moved down the hallway (see Figure 2.2). Turtlebot A was responsible for following the operator at a fixed distance and localizing the operator relative to itself using the Kinect sensor. Turtlebot B was responsible for patrolling in front of the operator as he moved down the hall. The human operator was equipped with a Wiimote (Figure 2.1b) to relay commands to the swarm. The operator's Wiimote could be used to start or stop the robots, or switch the behavior of the lead robot from patrolling to leading the operator by a fixed distance.

We successfully implemented the algorithm and demonstrated it on the second floor hallway in the Computer Science wing of the TMCB at BYU. This hallway was approximately 2.5m in width with small amount of clutter in the form of doorways and tables temporarily placed in the hallway. The human operator was able to advance down the hallway while the two robots maintained correct positions. Turtlebot A was able to keep itself and the operator localized relative to the map, and relay the information to Turtlebot B, which patrolled in front of the operator as he moved. The robots successfully responded to commands from the Wiimote during the experiment.

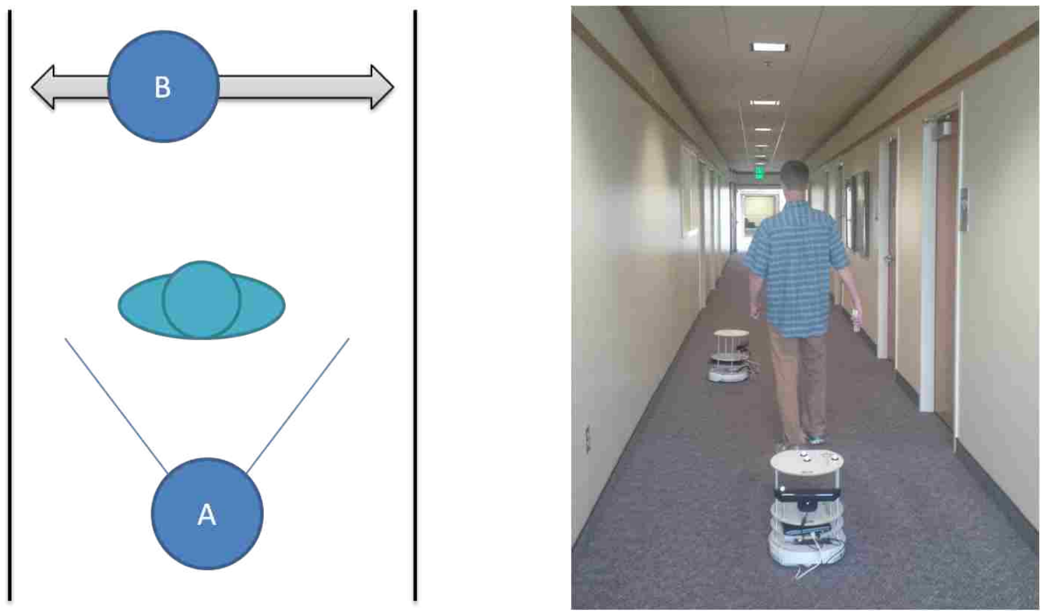


Figure 2.2: Two TurtleBots escort a human operator in a simulated force protection scenario. Robot A is responsible for tracking and localizing the human operator, while robot B is responsible for patrolling ahead of the operator. The operator is equipped with a Wiimote used to issue commands to the robots.

We then repeated the experiment with a remote operator by relaying position commands to the robots rather than using Turtlebot A to provide operator localization. The transmitted location was used in place of a proximate operator and the TurtleBots maintained position relative to this location. The remote human operator advanced this location down the hall and transmitted commands using the Wiimote. During the experiment, the TurtleBots maintained correct positions relative to transmitted location and responded to Wiimote commands.

2.3.3 Feasibility of Couzin's Model

In addition to validating HSI, we also validated the feasibility of Couzin's model on the TurtleBots. To allow the TurtleBot to navigate around the lab, it was necessary to introduce an obstacle avoidance term into the model. This was accomplished by redefining the bounding term b_i, to provide obstacle avoidance, rather than confining the robot to a specific area. The TurtleBot uses a dynamic occupancy grid that incorporates information gathered from

Parameter	Description	Value
s	Forward velocity	$0.2m/s$
θ	Turning rate	$90°/s$
α	Blind spot angle[5]	$0°(135°)$
R_{obs}	Obstacle avoidance radius	$0.8m$
R_r	Zone of repulsion radius	$0.8m$
R_o	Zone of orientation radius	$2.0m$
R_a	Zone of attraction radius	$4.0m$

Table 2.2: The parameters for Couzin's model used during validation. Using these parameters, a TurtleBot successfully navigated the BYU MAGICC lab using on-board sensors.

sensors as well as map data (if available). To introduce obstacle avoidance into Couzin's model, we first define a obstacle avoidance radius, R_{obs}, and react to all occupied cells within that range.

$$n_i^o = \{j : \|c_i - c_j^o\| \leq R_{obs}\} \tag{2.1}$$

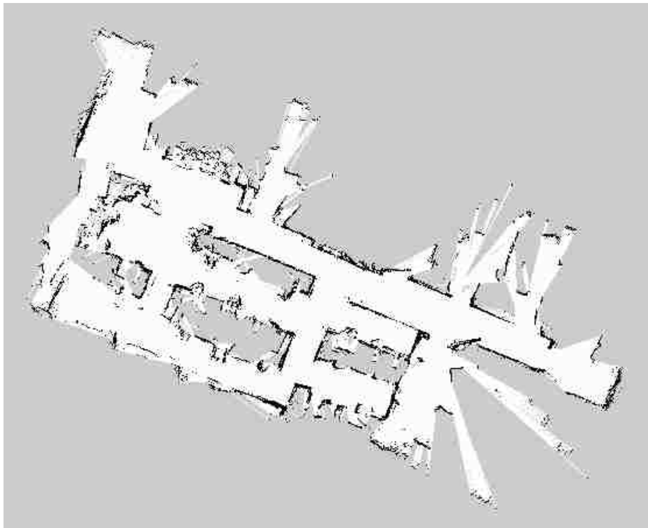
$$b_i = -\sum_{n_i^o} \frac{(c_j^o - c_i)}{|c_j^o - c_i|} \tag{2.2}$$

where c_j^o is location of each occupied cell j in the occupancy grid and b_i is used as described in Equation 1.15.

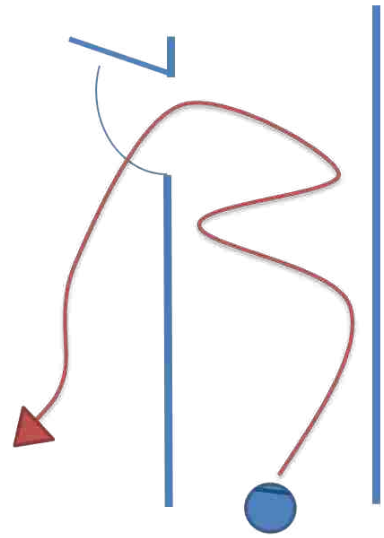
We then implemented this modified model on the TurtleBots and conducted an experiment to validate it. Our goals were to (1) test our obstacle avoidance modifications to Couzin's model, (2) verify that the hallways provided sufficient space for the TurtleBot to navigate and avoid obstacles, and (3) identify any real-world issues not found in our simulations. We conducted several runs in which we placed one of the TurtleBots in a hallway, initialized its approximate location, and allowed it to wander through the hallway. We found that the model was sufficient to allow the TurtleBot to move through the hallway in a random walk[6] and successfully avoid obstacles. Based on our experiments, we conclude that Couzin's

[5]We used $\alpha = 0$ during our test, allowing to robot to react to all obstacles it is aware of. The persistent occupancy grid largely compensated for the Kinect's 135° blind spot. See Sections 2.2.3 and 2.3.1.

[6]Note that in the absence of goals or additional robots, Couzin's model effectively produces a random walk as the robot moves forward and turns to avoid obstacles. The goal of the experiment was to verify



(a)



(b)

Figure 2.3: A SLAM-generated map of the BYU MAGICC lab. The MAGICC lab (a) consists of student workspaces, common areas, and hallways approximately 1.5-2m wide. During our validation of Couzin's model on the TurtleBots, our robot was able to navigate the hallways and avoid obstacles. Note that in the absence of goals or additional robots, Couzin's model effectively produces a random walk (b).

model is feasible and can be implemented on real robots provided (a) the robot can sense obstacles in its direction of travel and (b) the considerations in Section 2.2 are taken into account.

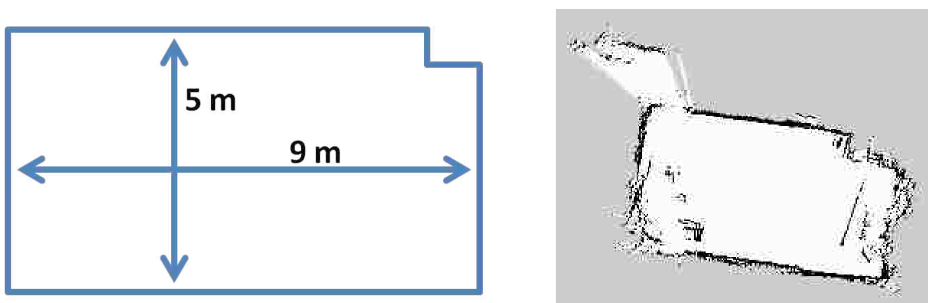
2.4 Modeling Robot Localization Error

Our experiments in Section 2.5 and Chapter 4 required an accurate model of our robots. This included verifying their physical parameters and determining an accurate model for the robots' localization error. Once the parameters were verified, we could then use the Stage robot simulator to conduct experiments on larger groups of robots and verify that the expected collective behaviors were present.

We used the BYU MAGICC lab's motion capture room (Figure 2.4), which is equipped with a camera system from Motion Analysis. The system provides position, orientation, and

that our modifications to Couzin's model were sufficient to allow navigation and obstacle avoidance. Later experiments with Stage validated flocking behaviors and group navigation (Section 2.5).

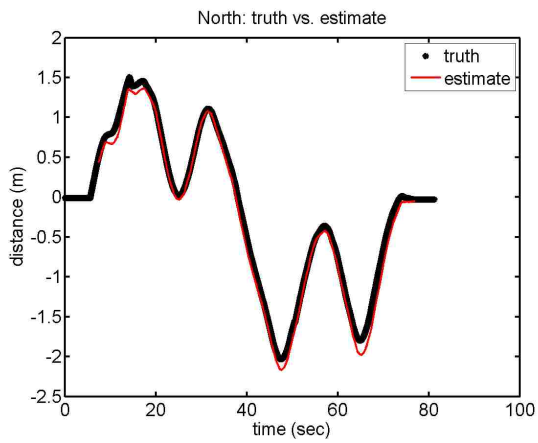
Figure 2.4: The MAGICC lab's motion capture room. The room is equipped with a camera system from Motion Analysis which provides 6DoF tracking with sub-millimeter and sub-degree accuracy at rates up to 200 Hz.



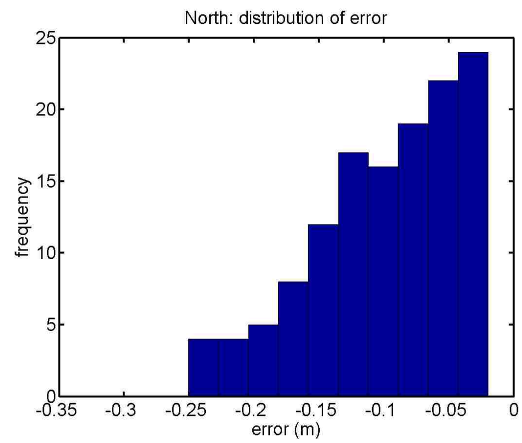
(a) (b)

Figure 2.5: Diagram of the motion capture room (a) and a SLAM map generated by TurtleBot. Note that the SLAM map incorporates furniture and other obstacles (Figure 2.4) in addition to the dimensions of the room.

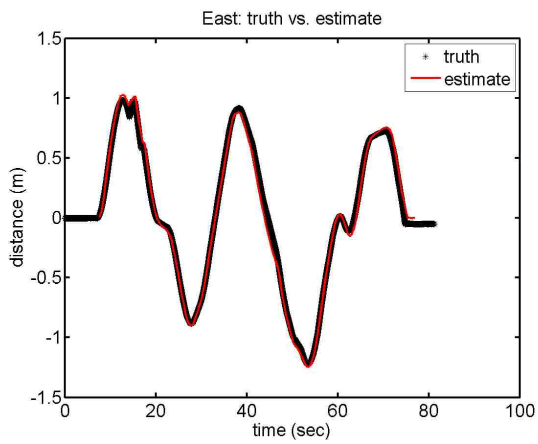
velocity information with sub-millimeter and sub-degree accuracy at rates up to 200 Hz. The TurtleBot was configured with markers and initialized to match the coordinate frame of the motion capture system. We then used the TurtleBot to create a SLAM map using the gmapping algorithm, with a grid size of 5cm (Figure 2.5). We found that the dimensions of the SLAM map match the measured size of the room to within a few inches. Finally, we teleoperated the TurtleBot around the room while recording data from both the camera system and the Turtlebot.



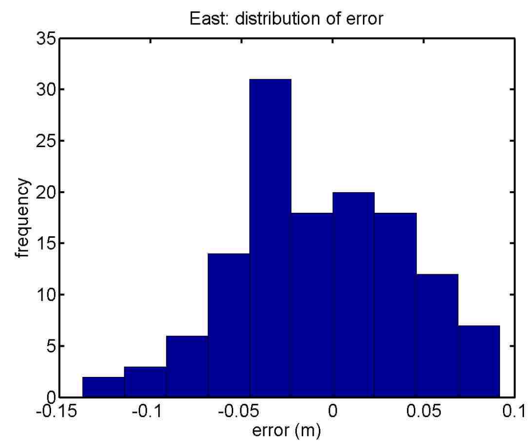
(a) North



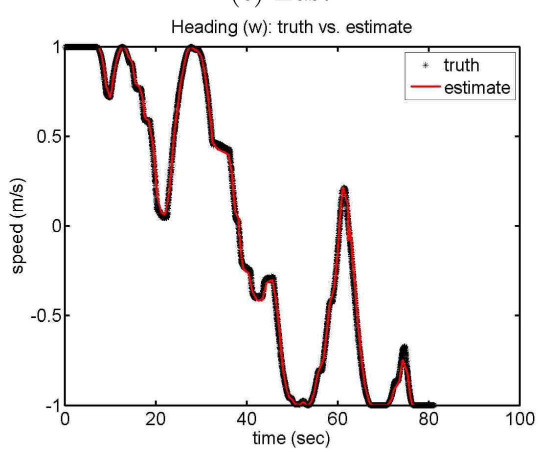
(b) North error



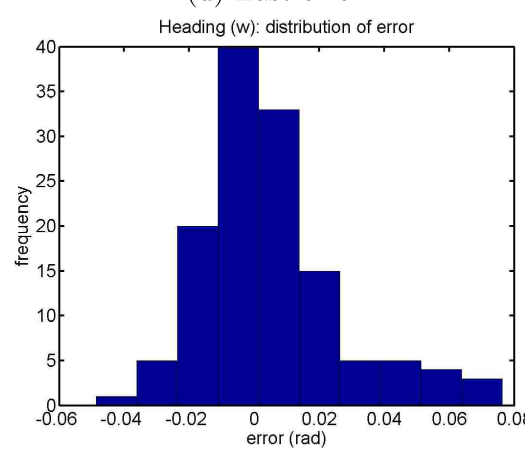
(c) East



(d) East error



(e) Heading (w)



(f) w error

Figure 2.6: True location vs the TurtleBot's estimate (a, c, e) and distribution of error (b, d, f). Truth values are shown in black, while the TurtleBot's estimate is shown in red. The TurtleBot's position and heading estimates trended well and closely matched the truth data obtained from the camera system. We found that the estimated heading was accurate to $\pm 1.2^\circ$ and the estimated position was accurate to $\pm 5cm$.

We then analyzed the robot's localization error by comparing the robot's estimate of its location to data gathered from the camera system. The state space of the robot is given by a position (north, east, down), and a quaternion orientation (x, y, z, w), giving $x = [n, e, d, x, y, z, w]^T$. As shown in Figure 2.6, the robot's estimate trends well and closely matches the true position given by the camera system. We then computed the absolute localization error (Figure 2.6). We found that the estimated heading was accurate to $\pm 1.2°$ and the estimated position was accurate to $\pm 5cm$, which was the granularity of our SLAM map. We also computed the RMS error (Equation 2.3) and covariance of the error (Equation 2.4) and found very little cross-correlation between error terms. Other parameters of the robot matched those in Table 2.1.

$$error_{RMS} = \begin{pmatrix} 0.1139 \\ 0.0467 \\ 0.4217 \\ 0.0049 \\ 0.0054 \\ 0.0254 \\ 0.0225 \end{pmatrix} \tag{2.3}$$

$$\sigma^2 = \begin{pmatrix} 0.0031 & 0.0003 & 0.0000 & -0.0000 & -0.0000 & 0.0003 & -0.0001 \\ 0.0003 & 0.0022 & 0.0000 & -0.0001 & -0.0000 & 0.0001 & -0.0003 \\ 0.0000 & 0.0000 & 0.0000 & -0.0000 & -0.0000 & 0.0000 & -0.0000 \\ -0.0000 & -0.0001 & -0.0000 & 0.0000 & 0.0000 & -0.0000 & -0.0000 \\ -0.0000 & -0.0000 & -0.0000 & 0.0000 & 0.0000 & -0.0000 & 0.0000 \\ 0.0003 & 0.0001 & 0.0000 & -0.0000 & -0.0000 & 0.0006 & 0.0000 \\ -0.0001 & -0.0003 & -0.0000 & -0.0000 & 0.0000 & 0.0000 & 0.0005 \end{pmatrix} \tag{2.4}$$

2.5 Validation of Flock Structures

One of the interesting properties of swarm algorithms is the collective behaviors that they display. Couzin's model displays several collective structures that are frequently found in schooling fish (see Figure 2.8). For our research, we wanted to validate that these collective

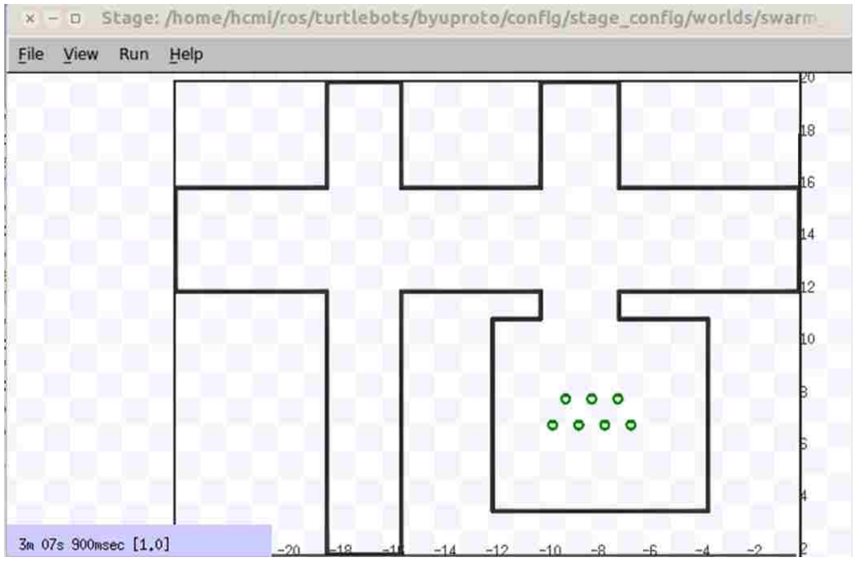
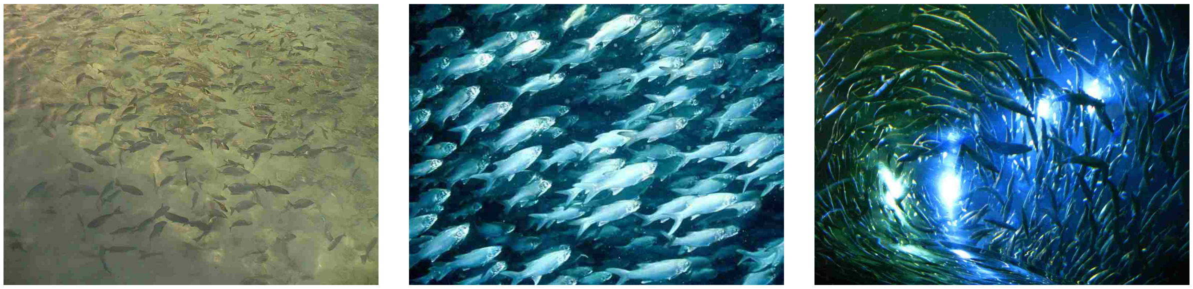


Figure 2.7: The Stage robot simulator. Using our data from Section 2.4, we modeled 7-8 TurtleBots and simulated them simultaneously.

behaviors could be reproduced on physical robots. Due to the limited number of robots at our disposal, we used the stage robot simulator to validate the existence of these collective behaviors. Stage is a 2D simulator that simulates the locomotion, sensors, and odometry of a variety of robots that are commonly used in research (see Figure 2.7). It can simulate multiple robots in real-time and allows sensor and odometry error to be simulated. Stage integrates with ROS to abstract robot hardware and allows the same code to be run on both simulated and physical robots without modification.

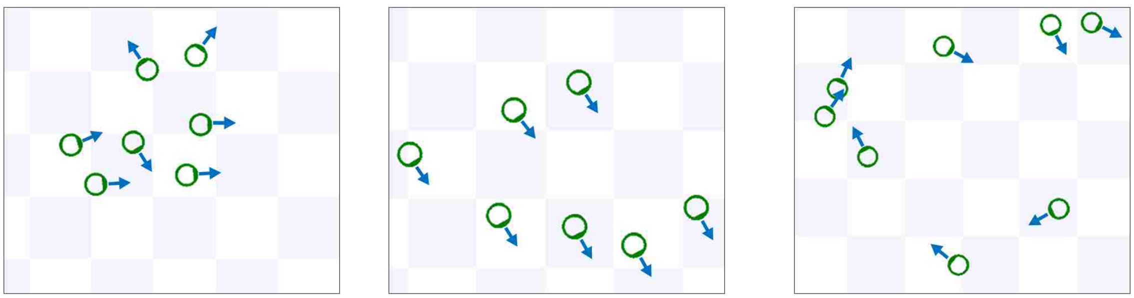
We used the physical parameters from the TurtleBots (Table 2.1), along with our own measurements (Section 2.4) to construct a TurtleBot model, including sensor and odometry error and configured the simulator to run 7-8 TurtleBots simultaneously. Because ROS provides an abstraction layer for robot hardware, each simulated robot ran the same software used in Section 2.3.3, with minor modifications to allow multiple copies to run on the same computer. As shown in Figure 2.9, we were able to reproduce all three flock structures on the simulated TurtleBots. Table 2.3 list the parameter values used in this experiment.

[7]Photo by Ibrahim Hussain Shihab and licensed under the Creative Commons Attribution Non-Commercial Share-Alike 2.0 license. The original photo can be found here: `http://www.flickr.com/photos/aindhy/3522319991/`.



(a) Swarm[7] (b) Flock[8] (c) Torus[9]

Figure 2.8: The collective behaviors of fish shown by Couzin's model.



(a) Swarm (b) Flock (c) Torus

Figure 2.9: The collective behaviors shown by the simulated TurtleBots. We were able to successfully replicate the collective structures shown by Couzin in simulation. Turtlebots are represented by circles, with an arrow to indicate the direction of travel.

		Behavior		
Parameter	Description	Swarm	Flock	Torus[10]
s	Forward velocity	$0.2m/s$	$0.2m/s$	$0.5m/s$
θ	Turning rate	$90°/s$	$90°/s$	$20°/s$
α	Blind spot angle	$0°$	$0°$	$0°$
R_{obs}	Obstacle avoidance radius	$0.5m$	$0.5m$	$1.6m$
R_r	Zone of repulsion radius	$1.0m$	$1.0m$	$0.8m$
R_o	Zone of orientation radius	$1.0m$	$2.0m$	$1.5m$
R_a	Zone of attraction radius	$5.0m$	$5.0m$	$5.0m$

Table 2.3: The parameters for Couzin's model used during our validation of collective structures on simulated TurtleBots.

2.6 Conclusion

The results presented in this chapter provide a concrete foundation and justification for the rest of the experiments found in this thesis. We successfully demonstrated that (1) an abstract algorithm can be implemented on real robots, without requiring additional information or supplemental hardware, (2) that these robots can be used in both remote and proximate HSI, and (3) the collective behaviors shown by Couzin can be reproduced on general purpose robots. Additionally, we showed that the sensing and processing capabilities of the robots were sufficient to run the algorithms, without requiring sophisticated sensors or advanced data processing, demonstrating feasibility on low-cost swarm robot hardware.

[8] Couzin distinguishes between *dynamic parallel groups* and *highly parallel groups*. However, in this thesis we consider them together under the heading of *flocks*.

[9] Photo by Flickr user Ewar Woowar `http://www.flickr.com/people/ewarwoowar/` and licensed under the Create Commons Attribution Non-Commercial Share-Alike 2.0 License. The original photo: `http://www.flickr.com/photos/ewarwoowar/1343993122/sizes/l/in/photostream/`.

[10] Theoretical work by Kerman [43] has shown that the radius of the torus formation is a function of forward velocity and turning rate. For a stable torus formation, the radius must be larger than $2R_r$, and we adjusted the turning rate accordingly. Other parameters were updated to compensate for the lower turning rate.

Chapter 3

Swarm Robot Systems: Effects and Considerations

3.1 Introduction

In this chapter we present our previously published work on swarm topology and performance drawn from two publications [33, 32] and a poster presentation [34]. This line of research primarily investigates the effect of different control methods and neighborhood definitions on swarm performance, the topological stability of the swarm, and interaction with human-controlled agents. The overall goal of this research was to understand the interactions and performance afforded by different swarm structures and control methods, and identify any tradeoffs. The findings presented here contribute to the formalization of HSI and provide practical results for real-world swarm implementation.

Some important challenges in the field of human-swarm interaction center on the topology of the swarm and changes in topology over time. One undesirable effect in swarm systems is *fragmentation* – the breakdown of the collective into fragments or individuals – which prevents the swarm from functioning as a cohesive unit. Another undesirable effect is *scrambling of the group*, which can be identified by tracking how fast the topology of the group changes. Scrambling within the group can reduce efficiency due to the extra effort required to maneuver through and around other agents. Additionally, scrambling can introduce sensor noise and reduce the effectiveness of certain sensors, such as UAV-mounted cameras, due to frequent changes in attitude. A third area of interest is the *expressiveness of the control method* used to influence the swarm, which we define as the ability to effect changes to the overall behavior and topology of the swarm (e.g. by splitting it into subgroups).

The overall group behavior and topology is built bottom-up from interactions among individuals and their environment [61]. Designing individual behaviors that produce desired collective behaviors is known as *swarm engineering* [71] and is still an open area of research. Consequently, it's important to understand the effect different design choices have on the collective behavior. In this research, we investigate the changes in collective topology and performance that result from different neighborhood definitions and control styles. Our results contribute to a better understanding of swarm systems and help to progress the field of swarm engineering.

This chapter presents the results from each paper sequentially in the order they were published. Section 3.2 presents our findings on the performance of parameter vs. predator control (poster presentation at AAMAS, 2011 [34]). Section 3.3 examines the effect of neighborhood definition and control style on the topology of the group and group interaction with the leader/predator (SMC, 2011 [33]). Section 3.4 reinforces the results from Section 3.3 with two small user studies (RSS, 2012 [32]). The results presented in this chapter were published in concert with other experiments and theoretical results from our research lab, and are best viewed in the context of their respective publications. The work presented in Chapter 4 and 5 is a direct continuation of this work, expanding the view to human operators and the individual robots in the swarm.

3.2 Parameter Control vs Predator Control, AAMAS 2011

This section contains our research presented at the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2011 [34]. The poster was primarily authored by Goodrich, with contributions from Pendleton, Sujit, Pinto, and Crandall. The author contributed simulations on the relative performance of parameter-based and predator-based control, comprising the section titled *Biomimetics*, which we present here.

3.2.1 Overview

The goal of this experiment was to investigate the tradeoffs in performance offered by bottom-up control vs top-down control (see [46]). We investigated the time to complete two information foraging tasks using two control strategies: (1) actively guiding the swarm with a predator, and (2) employing subjectively optimized set of parameters that instructs the swarm to spread out as much as possible. Parameter-based control performed better overall, while predator-based control showed interesting behaviors, such as the ability to split the group. The results from the experiment indicate a potential for the predator-based control to perform better on scouting tasks with a large quantity of information per location.

3.2.2 Experiment Description

We evaluated predator-based and parameter based-control using two variations of an information foraging task, which we refer to as "area coverage" (Figure 3.1) and "scouting" (Figure 3.2). The 120 x 120 playing field contains multiple points of interest, represented graphically by barrels, each containing a fixed quantity of information as determined by the experimental condition. The goal of the task is to deplete all of the barrels on the screen in as little time as possible by guiding 100 agents around the playing field. The barrels are depleted at a rate of 1 unit per second for each fish within $\sqrt{10}$ units of the barrel. The agents, represented graphically by fish, do not seek barrels on their own, but depend on the human operator to guide them to the targets. The two variations of the information foraging task are evaluated, each with a low-information and high-information condition, for a total of four scenarios. Table 3.1 gives a summary of the four experimental conditions. We evaluate predator-based and parameter-based control on each of the four scenarios using an "Oz of Wizard" style experiment[1] [66] run by the author.

[1]See Steinfield [66]. In "Wizard of Oz" experiments a real human participant interacts with a simulated system (e.g. an "autonomous" robot that is actually remote controlled by a hidden researcher). "Oz of Wizard" reverses these roles by simulating a human operator to a real system. Automated user interface testing in commercial software is an example of an "Oz of Wizard" scenario.

Scenario	Type	Information per Barrel
1	Area coverage	1
2	Area coverage	10
3	Scouting	100
4	Scouting	200

Table 3.1: Experimental conditions for the four scenarios presented in our AAMAS 2011 poster [34].

In this experiment, we used the augmented version of Couzin's model as described in Section 1.4.6 using the parameters listed in Table 3.2.

Area Coverage Task (Scenarios 1 and 2)

The first task, which refer to as "area coverage," consists of a uniform grid of barrels, each containing a small amount of information, which are placed 10 units apart over the playing field (see Figure 3.1). Scenarios 1 and 2 differ only in the amount of information contained in each barrel, set at 1 and 10 respectively.

Scouting Task (Scenarios 3 and 4)

The second task, which we refer to as "scouting" consists of 10 locations, each containing a relatively large amount of information, which are distributed randomly over the playing field (see Figure 3.2). Scenarios 3 and 4 differ only in the amount of information contained in each barrel, which is set at 100 and 200 respectively.

3.2.3 Results

Each scenario was run 5 times for each control method and the results are shown in Figure 3.3. We found that the subjectively optimized parameters performed better overall, but the predator-based control allowed for more flexible behaviors that could be adapted on-the-fly.

[2] Note that the special case $w_p = \infty$ induces a switching behavior, causing any agents within the radius of influence to ignore all agents except the leader or predator and those within their radius of repulsion. See Section 1.4.6.

		Control Method	
Parameter	Description	Predator	Parameter
s	Forward velocity	3.0	3.0
θ	Turning rate	$40°/s$	$40°/s$
α	Blind spot angle	45.0°	45.0°
R_r	Zone of repulsion radius	1.0	14.0
R_o	Zone of orientation radius	14.0	14.0
R_a	Zone of attraction radius	14.0	14.0
w_p	Leader/predator control gain[2]	∞	-
w_s	Stakeholders control gain	-	-
R_p	Leader/predator interaction radius	7.0	-

Table 3.2: The simulation parameters for predator-based and parameter-based control. All numbers are in simulation units unless otherwise noted. A dash ("-") indicates the parameter is not applicable to this experimental condition.

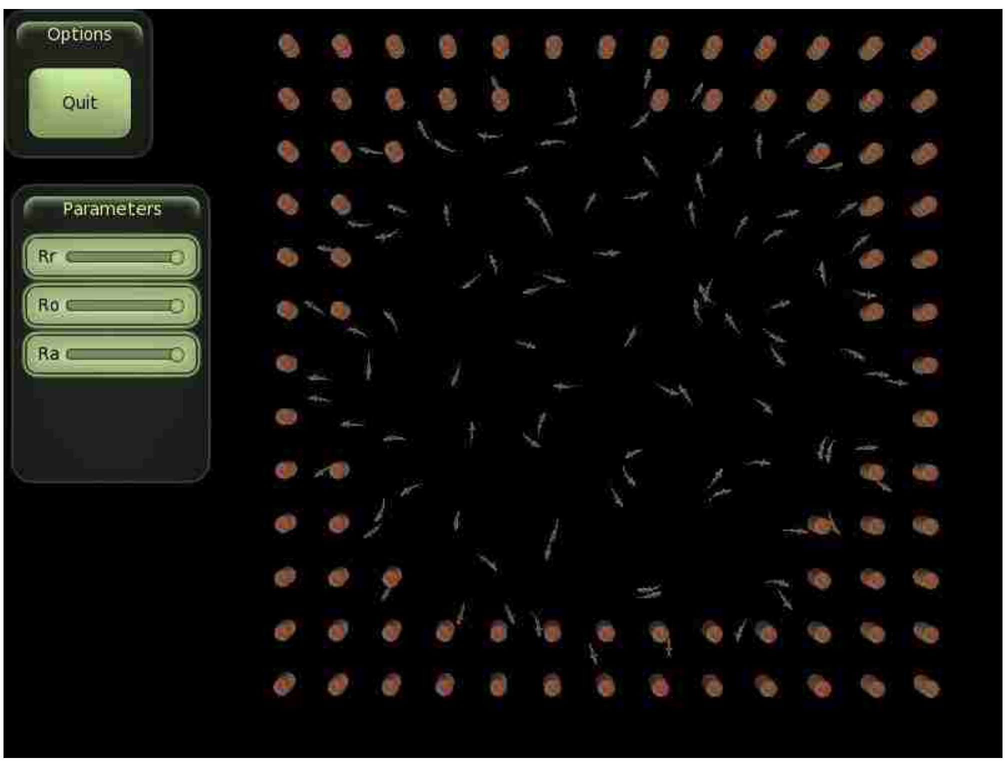
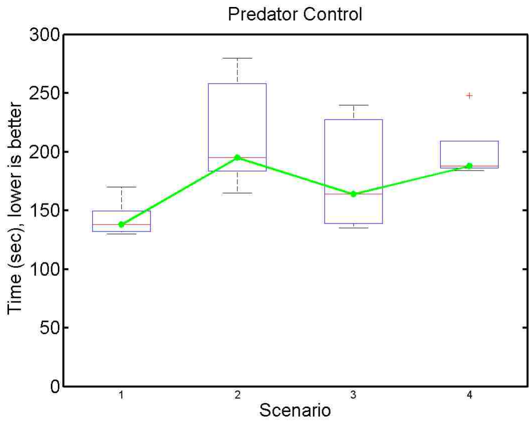


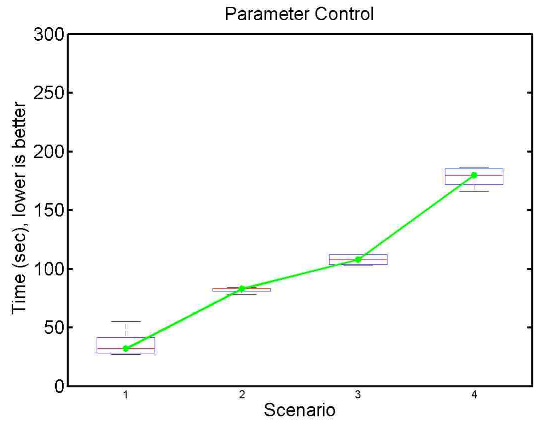
Figure 3.1: An area coverage task using parameter control. The agents must deplete a uniform grid of barrels, representing coverage of the entire playing field. The agents under parameter control are optimized to spread out and conver the area evenly.

Figure 3.2: A scouting task under predator control. The agents must deplete ten barrels randomly placed around the playing field.

Additionally, the two control methods completed the scenarios in different ways. Parameter-based control kept a small number of agents at each target for a long period of time, while predator-based control made passes with a large number of agents, depleting each barrel quickly in turn. Consequently, the performance of predator-based control stayed relatively flat, while performance with parameter-based control decreased as more information was added, indicating that predator-based control may perform better overall for sparse large-scale tasks where the time to complete the task is much greater than the travel time between tasks. Finally, predator-based control was capable of more expressive behaviors, such as splitting the group of agents into subgrounps, or guiding them along more complex paths and may be a better choice for some situations where the nature of the environment is not known apriori. This increased expressiveness came at the cost of additional control effort. The predator-based control needed to be actively managed, while the parameter based control did not.



(a) Predator control



(b) Parameter control

Figure 3.3: The time to complete the four scenarios under predator and parameter control (lower is better). Scenarios 1 and 2 are an area coverage task (Figure 3.1). Scenarios 3 and 4 are a scouting task (Figure 3.2). Parameter control did better over all, but predator control shows a good trend and more expressive behaviors.

The results of this research suggest that there are trade-offs between simplicity of management and the expressiveness of the control style. In swarm management, performance, expressiveness, and control effort are also influenced by the control style. Additional research is needed to further evaluate and quantify these trade-offs.

3.3 Neighborhoods and Interaction Style, IEEE SMC 2011

This section contains our research published in the IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2011 [33]. The paper was primarily authored by Goodrich, with contributions from Pendleton, Sujit and Pinto. I contributed simulations and analysis of the A and B matricies under two control methods and two neighborhood definitions, as well as contributing ideas to the formalisms presented in the paper. My simulation results are presented under the heading *Biomimetics* and the results accompanying that section. We present these results, accompanied by additional graphs and analysis, in this section.

3.3.1 Introduction

One of the goals of HSI research is to allow a human operator to predictably direct the swarm towards a desired state. Consequently, an understanding of the state of the swarm and how that state evolves over time provide valuable insight for the design and analysis of human-swarm systems. In this paper, Goodrich begins by providing a formalization of the evolution of the swarm's state for human interaction with bio-inspired robot teams (HuBIRT). This formalization is then used to develop tools for analysis, which we used in the subsequent experiments as well as in the remainder of this thesis. We therefore summarize several key concepts introduced by Goodrich, upon which I based my contribution to the paper.

In the paper, we begin by defining the state of the swarm x_t as a vector containing the state of each individual agent x_t^i and consider the evolution of the state of the swarm over time:

$$x_t = [x_t^1, x_t^2, \dots x_t^N]^T \tag{3.1}$$

$$x_{t+1} = f(x_t, u_t) \tag{3.2}$$

where u_t is external input to the system, and f is a nonlinear function representing the dynamics of the system.

With some loss of generality, we now separate this equation into state changes due to (1) the dynamics of the swarm model, and (2) external input to the system[3], represented by the functions f and g respectively:

$$x_{t+1} = f(x_t) + g(x_t, u_t) \tag{3.3}$$

Because the collective behavior of the swarm depends on local interactions between each agent and its neighbors, we further decompose Equation 3.3 to consider the evolution of

[3]In this thesis, we consider only human input to the system.

the state of each individual agent i due to interactions with external input u_t and interactions with all other agents $\neg i$. This is accomplished by considering the ith row of Equation 3.3, representing the evolution of agent i's state, and partitioning x_t into the state of agent i, x_t^i, and the state of all other agents, $x_t^{\neg i}$:

$$x_{t+1}^i = f^i(x_t^i, x_t^{\neg i}) + g^i(x_t^i, u_t) \tag{3.4}$$

This formulation allows us to consider the effect of changes to swarm dynamics, f, and control methods, g, on the evolution of the state of the swarm over time. Specifically, we are interested in monitoring (a) the amount of scrambling within the group, and (b) the degree of influence sustained by the leader or predator. To facilitate this analysis, we define two adjacency matricies which represent (1) interactions among individual members of the swarm (NxN) and (2) interactions between external inputs u_t and the individual agents ($Nx1$). We refer to matrix (1) as the *cohesiveness matrix*, A_t, and matrix (2) as the *management matrix*, B_t, which are defined as follows:

$$A_t = \{1 : a_t^{ij} = 1\} \tag{3.5}$$

$$B_t = \{1 : b_t^i = 1\} \tag{3.6}$$

where a_t^{ij} means that agent i and j are interacting and b_i^t means agent i is influenced by the external input to the system u_t. The exact definition of interaction varies by model, but for this thesis we define:

$$a_t^{ij} = \begin{cases} 1 & \text{if } ||c_i - c_j|| < R_a \text{ and } |\psi^{ij}| < 180° - \alpha \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

$$b_t^i = \begin{cases} 1 & \text{if } ||c_i - c_p|| < R_p \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

where ψ_{ij} is the angle between agent i's heading (v_i) and the relative angle to agent j ($c_j - c_i$). Note that $b_t^i = 1$ indicates that the agent is close to the leader or predator and therefore under its influence. Additionally, the use of a switching controller ($w_p = \infty$) means that the agent is exclusively influenced by the leader or predator when it is in range and the other agents when it is not.

As noted in Section 3.1, changes in swarm topology and human influence have important real-world implications. In this section, we examine the effect of two different control styles, leader and predator, along with two different neighborhood definitions on the evolution of the swarm topology, A_t, and human influence, B_t, over time. Specifically, we evaluate how well the operator-controlled agents sustain influence (as evaluated by a small number of changes in the B matrix), and the stability of interactions within the swarm (as evaluated by a small number of changes in the A matrix). The goal of this research was to provide data that could be used in swarm design to minimize undesirable effects, such as scrambling within the group.

3.3.2 Experiment Description

We conducted an "Oz of Wizard"[4] style experiment [66] in which we simulated typical and worst-case interactions between the swarm and a human operator. The typical case consists of guiding the swarm in a "scouting" information foraging task as described in Section 3.2.2. The worst-case interaction is modeled by actively moving the leader or predator through the center of the group several times in a zig-zag pattern. This path was chosen in order to scramble the swarm as much as possible (maximizing changes to the A matrix), and spread the influence of the leader or predator (maximizing changes to the B matrix). The goal of the experiment was to learn which control style and neighborhood definition was most stable under typical and worst-case interactions.

[4] See footnote in Section 3.2.2 for a description.

Parameter	Description	Value
s	Forward velocity	3.0
θ	Turning rate	$40°/s$
α	Blind spot angle	45.0°
R_r	Zone of repulsion radius	1.0
R_o	Zone of orientation radius	14.0
R_a	Zone of attraction radius	14.0
w_p	Leader/predator control gain	∞
w_s	Stakeholders control gain	-
R_p	Leader/predator interaction radius	30.0

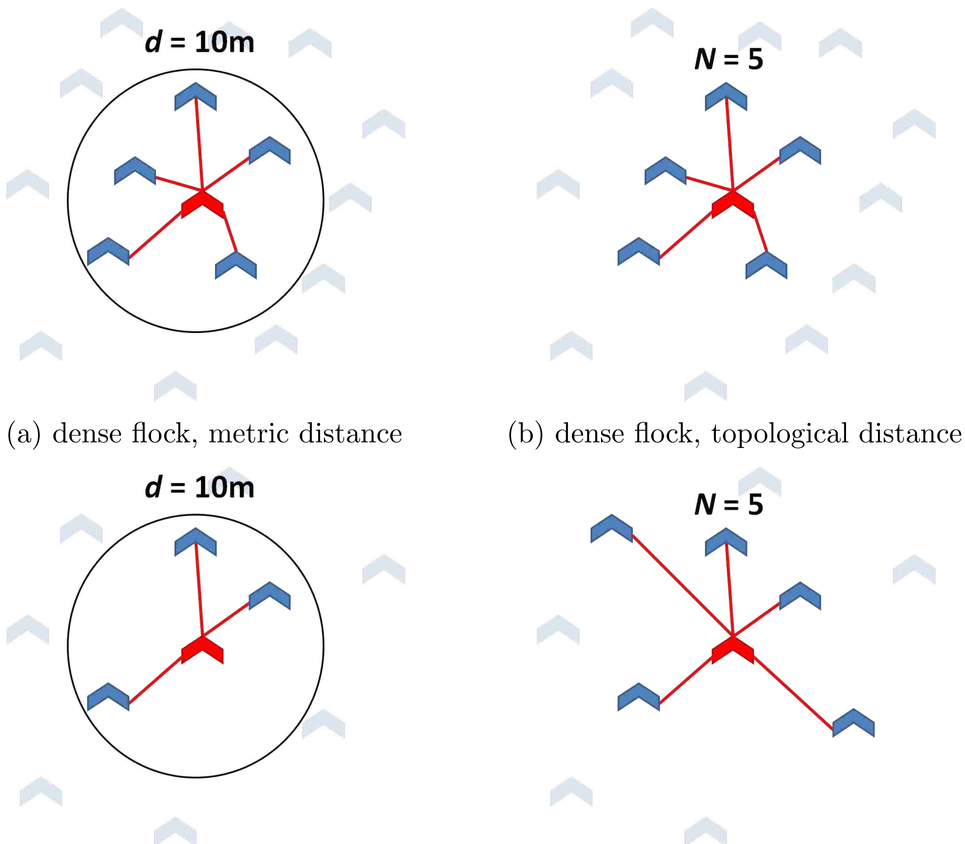
Table 3.3: The simulation parameters for predator-based and leader-based control. All numbers are in simulation units unless otherwise noted. A dash ("-") indicates the parameter is not applicable to this experimental condition.

We selected two control styles, leader and predator, and two neighborhood definitions, topological and metric, and evaluated the typical and worst-case interactions on each, resulting in a 2 x 2 x 2 experiment with eight conditions total.

Neighborhood Definitions Based on Topological or Metric Distance

In this experiment, we were inspired by the work of Ballerini et al. [7] on starling flocks. In their paper, they show that flocking starlings react to the nine closest birds, rather than reacting to all birds within a certain distance. In contrast, many existing flocking algorithms ([23, 55] for example) define a fixed interaction distance and react to all other agents within that range. We wanted to evaluate whether a neighborhood definition based on a fixed number of neighbors, instead of a fixed distance, would make the topology of the swarm more stable.

In this thesis, we define a *topological distance of* N as the N^{th} closest agent to a given agent. We then define a *topological neighborhood definition*, where agents react to all agents within a topological distance of N, and a *metric neighborhood definition*, where agents react to all agents within a fixed distance of d. Figure 3.4 provides a visual description.



(a) dense flock, metric distance (b) dense flock, topological distance

(c) sparse flock, metric distance (d) sparse flock, topological distance

Figure 3.4: The difference in neighborhoods in a sparse and dense flock using topological and metric distances. Using *metric distance* (a and c), an agent interacts with a variable number of other agents, depending on the density of the flock. Using *topological distance* (b and d), agents interact with a fixed number of neighbors regardless of flock density.

3.3.3 Analysis

To analyze the results, we examined the differences in the A and B matrices between consecutive time steps, and computed a time series representing the total number of changes at each step (Figure 3.6 and 3.9). We chose to compute the power spectral density[5] (PSD) of the time series rather than taking a simple average[6]. We hypothesized that leader and predator would differ, not only in the average number of changes, but in the way those changes occurred. A PSD plot allowed us to visualize not only the average changes but also changes that happened in bursts at regular intervals.

Additionally, we computed an interaction histogram for each agent and the leader or predator. We used the histogram to view the spread of interactions across agents. We wanted to know if a given leadership style or neighborhood definition created consistent or sustained interaction between individuals. This has importance not only for group stability but also for other topics, such as the construction of mesh networks among the agents in the swarm.

3.3.4 Results

We found that the leader sustained influence better than the predator, as shown by the smaller power spectrum in Figure 3.7. However, we found that because of this sustained influence, the leader caused more scrambling within the swarm (see Figure 3.10). The neighborhood definition, topological or metric, did not have a strong influence on the influence sustained by the leader or predator. However, the neighborhood definition had a strong effect on the stability of the group topology (see Figure 3.10). We also found that the predator interacted with fewer agents overall (see Figure 3.5).

The next two sections present a more detailed overview of our results.

[5] A power spectral density plot represents the power in each frequency component of a given signal.

[6] In the A matrix, each time series was computed per agent. The power spectral density was also computed per agent and then averaged across agents. This was done to avoid filtering out changes that occurred at a higher frequency.

3.3.5 Sustained Leader / Predator Influence (B Matrix)

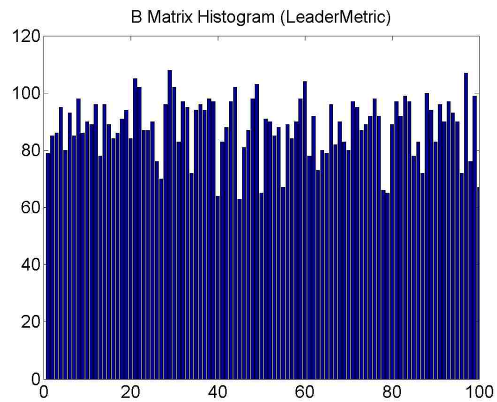
To determine which control method and neighborhood definition sustained influence better, we primarily examined the four "typical interaction" scenarios. The "worst-case interaction" scenarios are designed to disrupt the swarm by intentionally avoiding sustained influence and do not provide additional insight. We omit them for brevity, noting that the results are similar but less pronounced.

We found that leader-based control sustained influence better than predator-based control, as shown by the power spectral density plots in Figure 3.7. We also found that predator-based control interacted with fewer agents overall, and interacted with those agents for shorter time periods (see Figure 3.5). The neighborhood definition did not have a substantial impact on the sustained influence of the leader or predator.

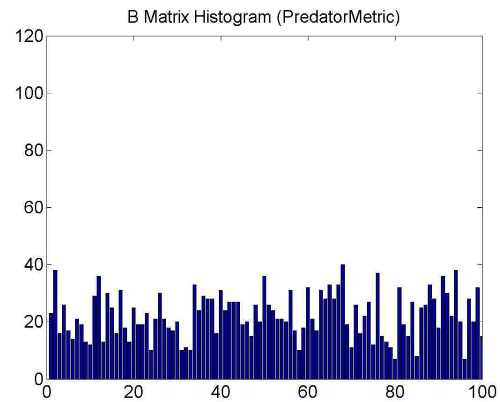
We hypothesize that the leader *affords* sustained influence by drawing interacting agents towards itself. The dynamics of the swarm model ensure that other agents follow the interacting agents and that influence is sustained as the leader steers the swarm. Similarly, the predator *affords* brief or impulse interaction. The repulsion of the predator causes interacting agents to turn toward the control vector u_i^p, drawn from the predator to the agent. Consequently, agents stop receiving the predator's influence once they act on it, which creates short "impulse" interactions. These ideas are further reinforced by the participant comments in Section 3.4.3 and our findings in Chapter 5. A full analysis of the affordances of these control primitives should be addressed in future work.

3.3.6 Topological Stability of the Group (A Matrix)

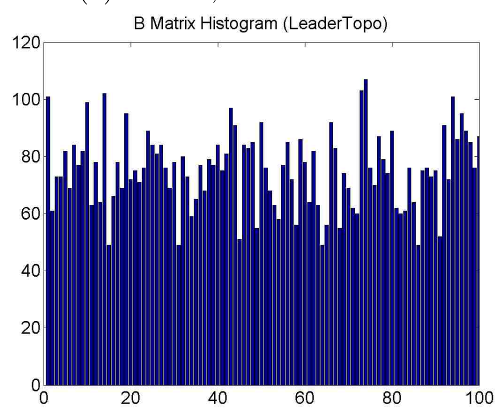
In this section we focus on the results provided by the "worst-case interaction" scenarios. Because the goal of these scenarios was to actively disrupt the swarm, they provide insight into the topological stability provided by the two neighborhood definitions and interaction styles. The results of the "typical interaction" scenarios are similar, but we omit them for brevity.



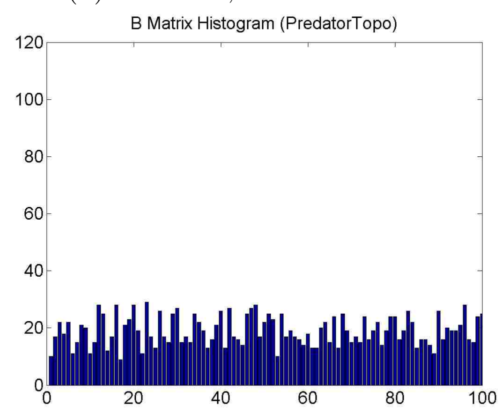
(a) Leader, metric distance



(b) Predator, metric distance

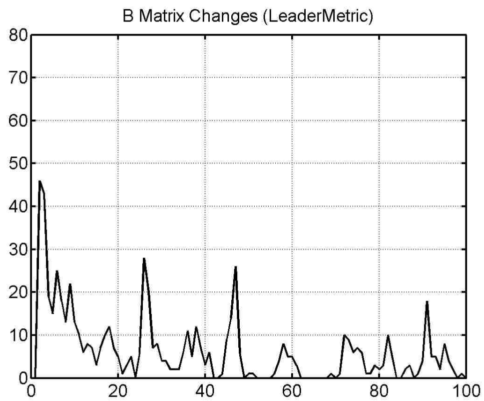


(c) Leader, topological distance

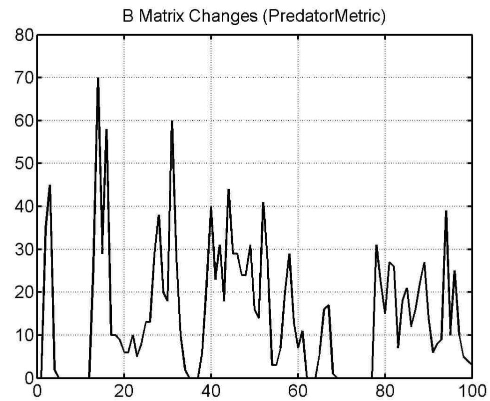


(d) Predator, topological distance

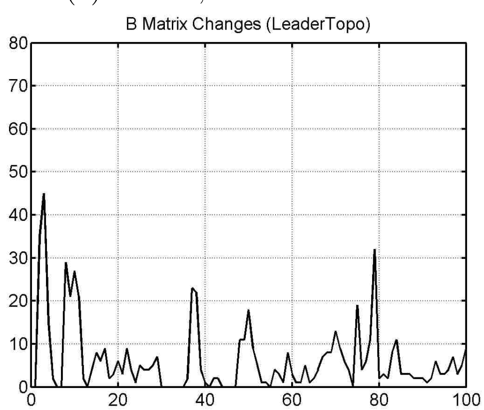
Figure 3.5: Histogram of the B matrix for the four "typical interaction" experimental conditions. The graph shows the total number of interactions with each agent during the scenario. Each bar represents one agent and the height of the bar shows the number of interactions. The predator interacted with fewer agents than the leader and interacted with those agents over shorter durations.



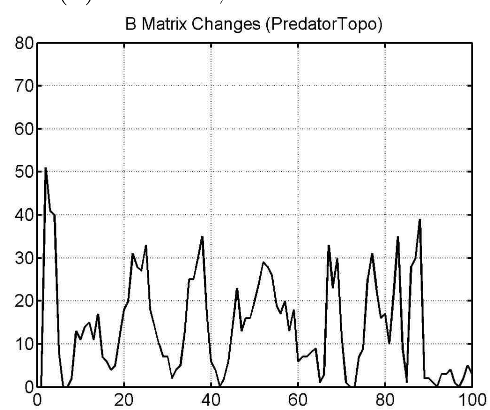
(a) Leader, metric distance



(b) Predator, metric distance



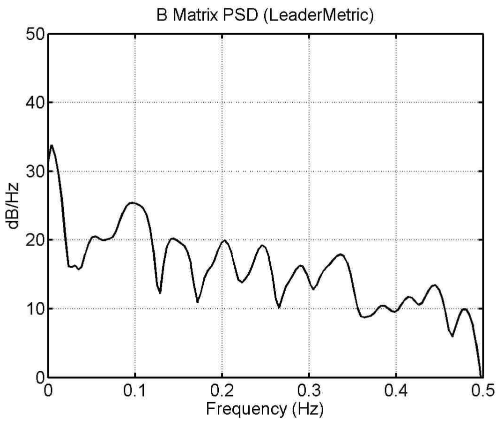
(c) Leader, topological distance



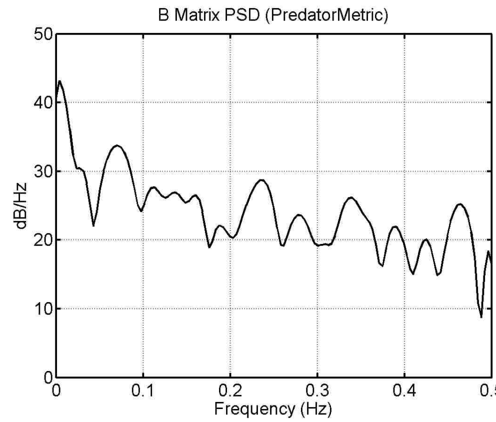
(d) Predator, topological distance

Figure 3.6: The number of changes to the B matrix at each time step for the four "typical interaction" experimental conditions. To analyze sustained influence, we created a time series that shows the number of entries in B_t that have changed since the previous time step, allowing us to visualize the magnitude and timing of changes to the B matrix. The size and frequency of the peaks show how well human influence is sustained (smaller = more sustained). We then use this time series to compute the PSD plots shown in Figure 3.7

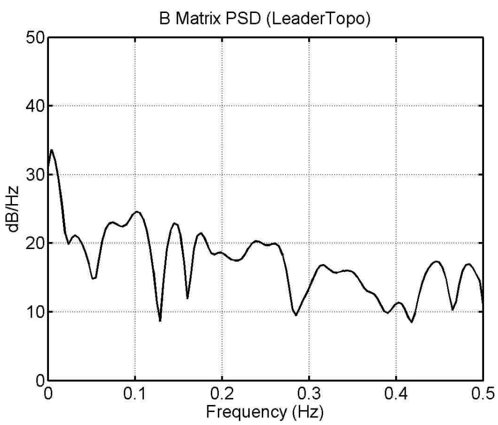
.



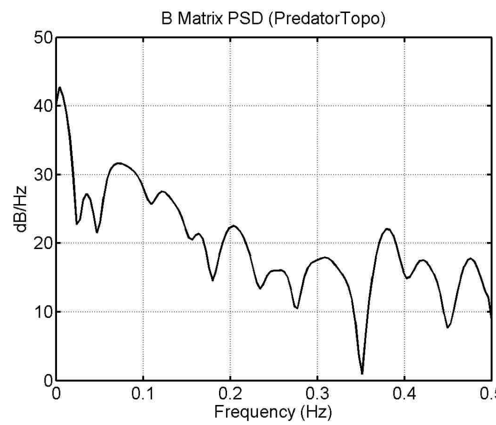
(a) Leader, metric distance



(b) Predator, metric distance

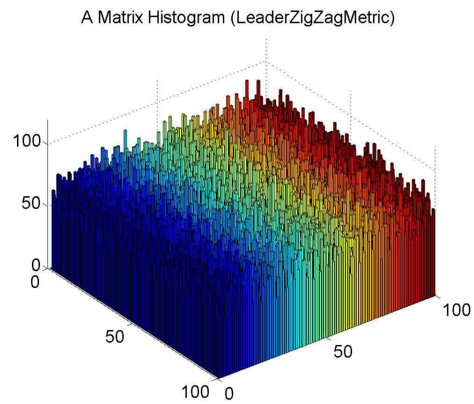


(c) Leader, topological distance

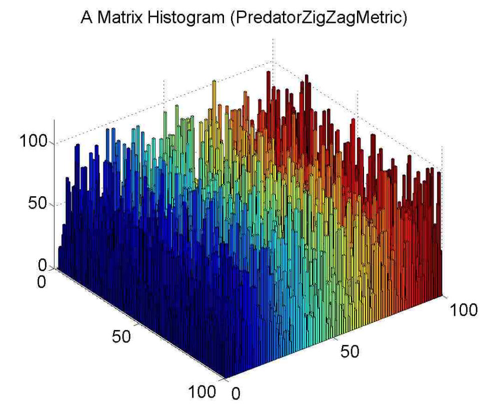


(d) Predator, topological distance

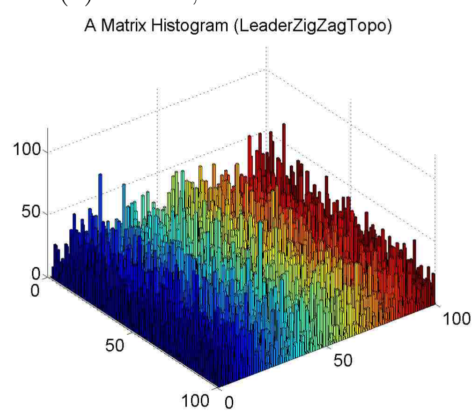
Figure 3.7: The power spectral density (PSD) of changes to the B matrix for the four "typical interaction" experimental conditions. The leader sustains influence better than the predator as shown by a smaller power over the frequency spectrum. Topological distance resulted in fewer high-frequency changes with predator control, but did not make a substantial difference overall.



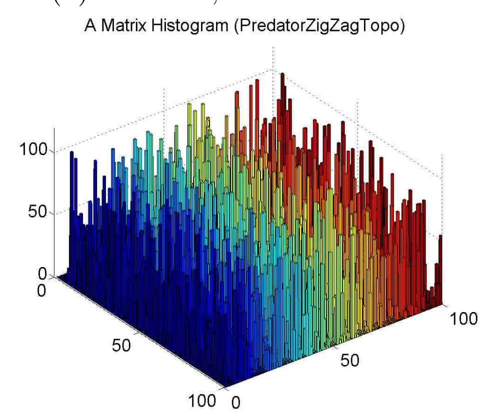
(a) Leader, metric distance



(b) Predator, metric distance

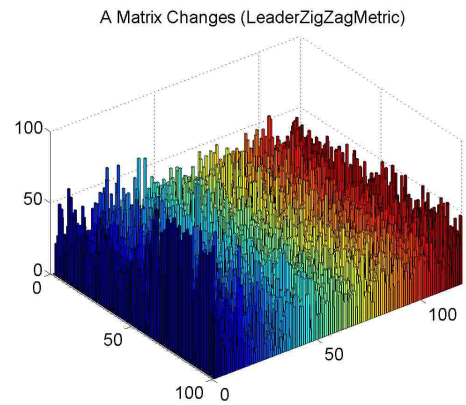


(c) Leader, topological distance

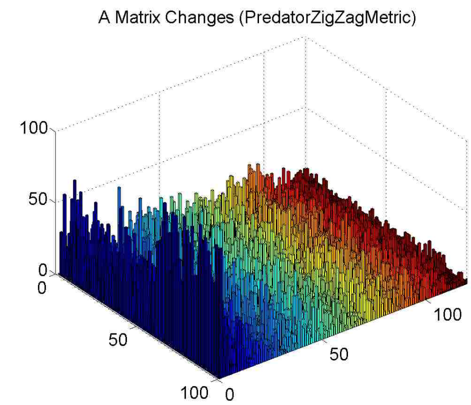


(d) Predator, topological distance

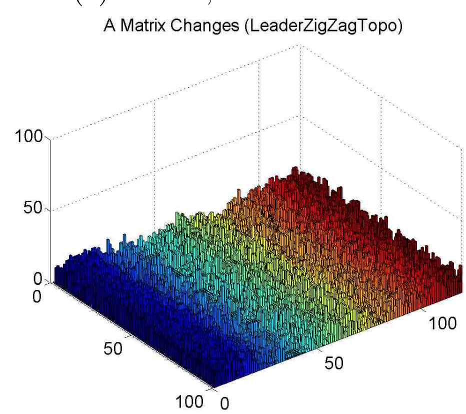
Figure 3.8: Histogram of the A matrix for the four "worst-case" experimental conditions where rows and columns are agents, and height is the number of times agent *row* and *column* interacted. The leader pulled the swarm closer together, resulting in more communications overall, and spread the interactions out more between agents due to scrambling within the group (also see Figure 3.10). Using topological distance, rather than metric distance, resulted in fewer agent-to-agent interactions under leader control.



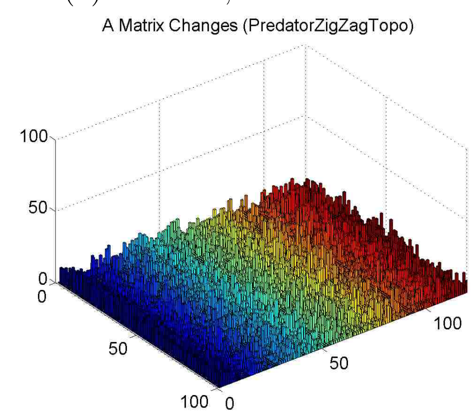
(a) Leader, metric distance



(b) Predator, metric distance

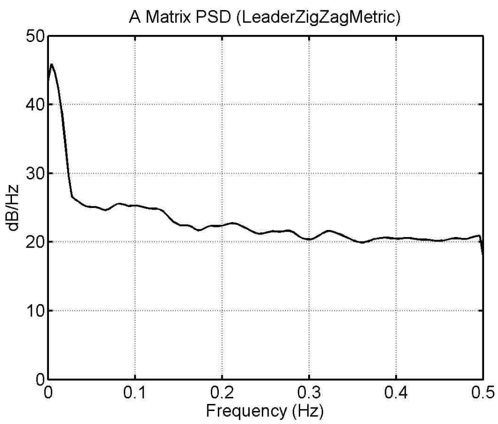


(c) Leader, topological distance

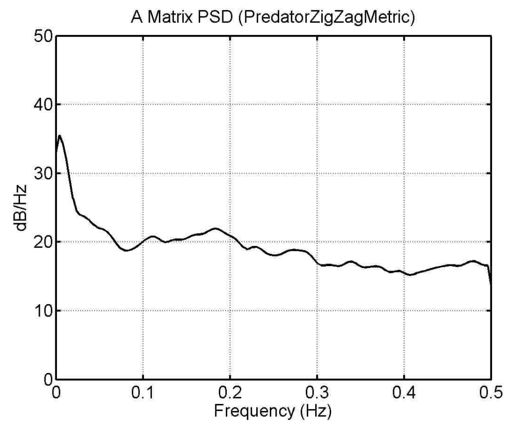


(d) Predator, topological distance

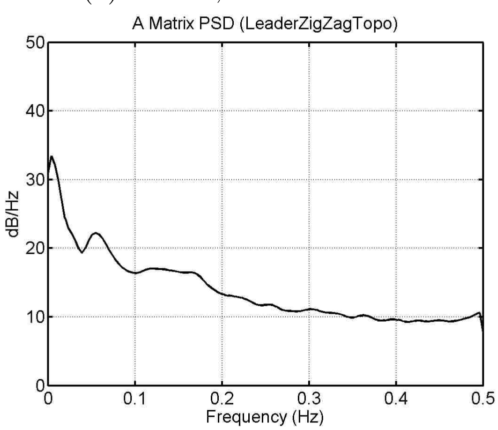
Figure 3.9: The number of changes to the A matrix at each time step for the four "worst-case interaction" experimental conditions where rows are agents, columns are timesteps, and height is the number of changes. Using topological distance, rather than metric distance, greatly increased the stability of the swarm topology (fewer changes). Predator control also caused fewer changes in group topology than leader control. Further analysis with power spectral density plots is shown in Figure 3.10



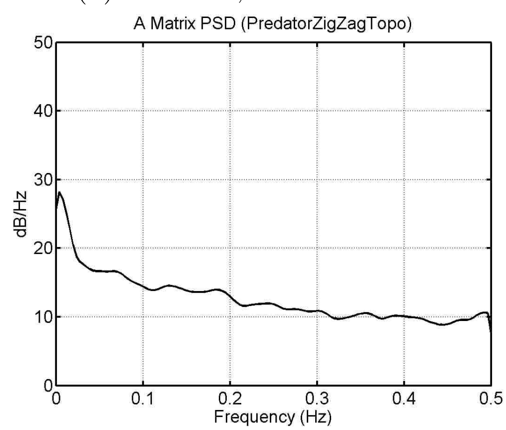
(a) Leader, metric distance



(b) Predator, metric distance



(c) Leader, topological distance



(d) Predator, topological distance

Figure 3.10: The power spectral density (PSD) of changes to the A matrix for the four "worst-case interaction" experimental conditions. Using topological distance results in fewer changes in group topology as shown by the smaller power spectrum. Additionally, the leader control caused more changes to group topology than predator control for worst-case interaction (see Figure 3.12 for the typical case).

We found that the topological neighborhood definition substantially reduced the number of changes to the topology of the swarm, as shown by Figure 3.10. We also found that predator-based control caused fewer changes than leader-based control to the topological structure of the group. We hypothesize that this is an indirect result of the switching behavior of the leader and predator ($w_p = \infty$). While the leader or predator is sustaining influence, the agents closest to the leader or predator bounce between human influence and repulsion from other agents. Consequently, because the leader sustains influence better than the predator, it also induces more scrambling of the swarm. This secondary effect is addressed in Chapter 5 where we use a non-switching control gain for the leader and predator.

3.4 Interaction Style User Study, RSS 2012

This section contains our research published in the Robotics Science and Systems Conference (RSS), 2012 [32]. The paper was primarily authored by Goodrich, with contributions from Pendleton, Kerman, and Sujit. The author designed, implemented, and analyzed two small-scale user studies (referred to as case studies) in addition to the results from Section 3.3, which are also referenced. The author's contribution is primarily found under the heading *Switching Controllers*, but also includes ideas contributed toward the formalisms found in the paper. Here we present the design and motivation for the case studies, along with graphs and analysis.

3.4.1 Introduction

Based on our results from the case study in Section 3.3.5, we decided to run a small-scale user study to see if the control style results held for human operators executing a goal-directed task. We primarily wanted to evaluate which leadership style better sustained influence during the study scenario.

3.4.2 Experiment Description

The experiment consisted of a "scouting" information foraging scenario as described in Section 3.2.2. The amount of food in each barrel was set at 5.0, with two barrels present initially and additional barrels appearing randomly as the scenario progressed. The model parameters used in the study were the same as those found in Table 3.3.

During each scenario, the position of all swarm members, the position of the leader or predator, and the current A and B matrices were logged at a rate of 1 Hz and saved to disk for later analysis. Each scenario was run for 120 seconds.

Each participant was given a handout with instructions for the experiment and control of the leader or predator and were instructed to gather as much food as possible by guiding the agents to the barrels with the leader or predator. They then completed an initial training scenario with each control method. They were told to practice until they felt comfortable and to ask any questions about the experiment or control methods. After the training, each participant completed two runs with each control method and then completed a short survey. The survey asked participants to comment on (1) any strategies they employed for each control method, (2) their preferred control method, and (3) their relative confidence between the two control methods. Additionally, the second user study also asked participants to describe strategies they employed for each size, and to describe their relative confidence between the two group sizes.

3.4.3 Results

We ran the initial user study as described and then conducted another follow-on study due to an unexpected confounding factor. During the first user study, several participants developed a novel control strategy using the predator, which involved splitting the swarm into several small groups and moving the predator rapidly between them. These results were interesting, but did not meet our initial aim of measuring the sustained influence of the leader or predator.

Consequently, we conducted a second user study and instructed participants to keep the group together as much as possible.

Overall, the leader sustained influence better than the predator (as shown in the second user study), but the predator was more expressive and allowed for novel strategies such as the one mentioned above. However, the increased expressiveness came at the cost of fragmentation of the swarm. None of the participants were able to keep all 100 agents together using the predator, even when instructed to do so.

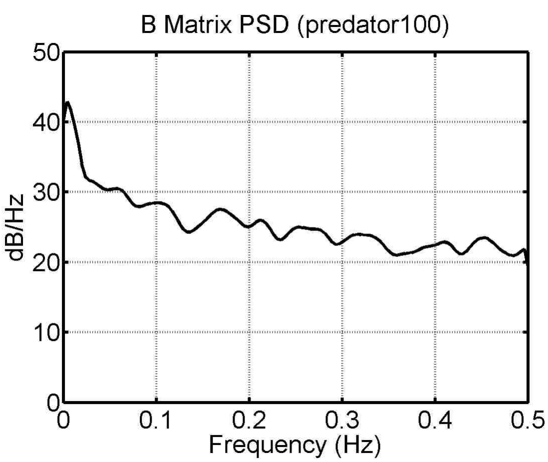
To analyze the results, we looked at the differences in the A and B matrices at each time step and computed a time series representing the total number of changes at each step. We chose to compute the power spectral density (PSD) of the time series rather than taking a simple average using the same method in Section 3.3.3. We hypothesized that leader and predator would differ, not only in the average number of changes but also in the way those changes occurred. A PSD plot allowed us to visualize not only the average changes but also changes that happened in bursts at regular intervals.

3.4.4 Initial User Study

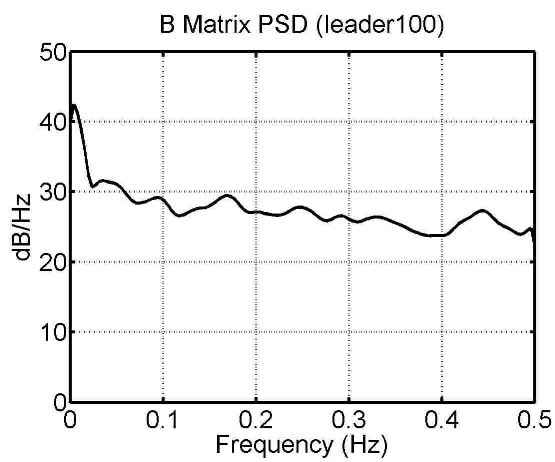
We recruited a convenience sample of six student volunteers from our research lab and obtained IRB informed consent. All six participants successfully completed the user study and no side effects were reported.

As mentioned above, several study participants employed a novel strategy for predator-based control. While it did not meet our initial aim for the experiment, it demonstrates a novel predator-based control strategy and the results provide some insight. First, it reinforces previous results that suggest that predator-based control may be more expressive than leader-based control, as demonstrated by the novel control strategy. Second, participants observed that they could gather more information by splitting the swarm into many subgroups, inadvertently replicating the results found in Section 3.2.3. Third, several participants mentioned that they thought the predator was more powerful, which reinforces the idea that the predator can be more expressive.

Because some participants split the swarm into small subgroups we omit the A matrix results but include the B matrix results for reference, noting that leader and predator used different control strategies. As shown in Figure 3.11, the average amount of change for the two strategies was similar, but the leader caused more changes at higher frequencies. This is probably due to the high speed at which some participants moved the predator when employing the novel control strategy mentioned above. The predator simply didn't stick around long enough to cause any higher frequency noise within the group. These results may indicate that a quick "impulse" control input may be better than sustained influence in some circumstances. However, the data gathered here is insufficient and we leave a full investigation of this point to future work.



(a) Predator



(b) Leader

Figure 3.11: The power spectral density (PSD) plots of the changes to the B matrix under leader and predator control. Both control methods have a similar number of average changes, but the leader has more changes at high frequency. We hypothesize that the short duration the predator interacted with any given agent did not provide an opportunity for it to introduce higher-frequency noise. This may indicate that short impulses of control may be better in some circumstances. However, this is left to future work.

3.4.5 Follow-on User Study

We recruited a convenience sample of five student volunteers from our research lab and obtained IRB informed consent. All six participants successfully completed the user study and no side effects were reported.

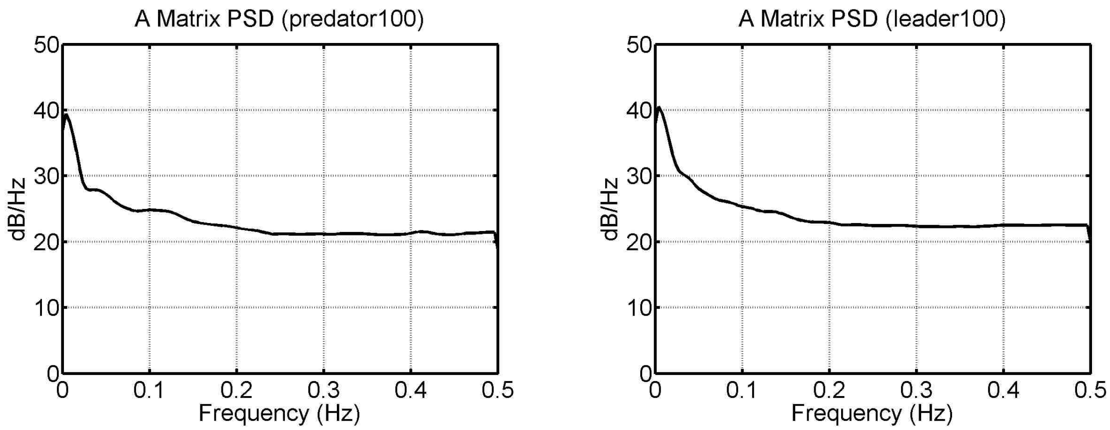
To avoid the issues in the previous study, we instructed participants to keep the agents together as much as possible while amassing information. The total amount of food gathered was logged to provide a measure of swarm performance. Additionally, we added a “small swarm“ and “large swarm“ condition to each of the control styles, which used 15 and 100 agents respectively. This brought the total number of scenarios up to four: two influence styles (leader and predator) and two swarm sizes (15 and 100 agents).

3.4.6 Performance

The total amount of food gathered was higher for the leader than for the predator, but only substantially so when using 15 agents. Study participants rated their confidence lower for 15 agents, especially when using the predator (see Table 3.4). Additionally, they mentioned that the swarm of 15 agents covered a smaller area as it moved and made it more difficult to hit the target locations, especially when using the predator. This suggests that the predator, while more expressive overall, was not as useful for making fine adjustments to the trajectory of the swarm. Future work should provide a more concrete investigation.

	Number of Agents	
Method	15	100
Leader	47.94 ± 8.40	53.77 ± 7.28
Predator	42.15 ± 6.00	53.67 ± 8.214

Table 3.4: Total information gathered



(a) Predator control (b) Leader control

Figure 3.12: Power spectral density PSD) of the A matrix. The leader induces slightly more scrambling than the predator, due to the switching behavior of the leader and predator ($w_p = \infty$) and the more sustained influence of the leader. This secondary effect is addressed in Chapter 5. Note that the graph uses a log scale (dB).

3.4.7 Topological stability of the group (A Matrix)

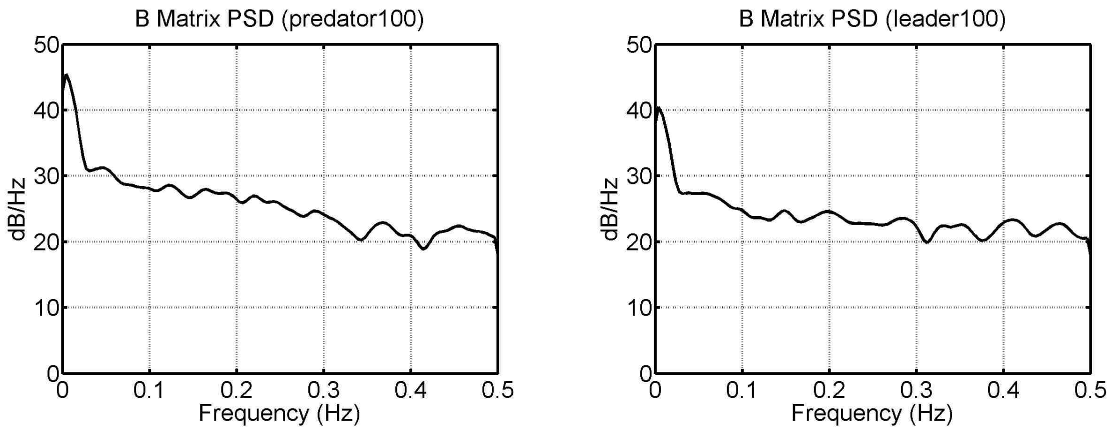
As shown in Figure 3.12, leader control caused slightly more scrambling within the group than predator control across all frequencies. It's important to note that the PSD plots are on a log scale (dB), so the small difference in the graphs is meaningful. These results replicate and reinforce our findings in Section 3.3.4.

3.4.8 Sustained Leader / Predator Influence (B Matrix)

As shown in Figure 3.13, leader-based control sustained influence better than predator-based control, as evidenced by the lower power in the PSD plot across all frequencies. Additionally, the shape of the power spectrum is different, indicating that the predator's B matrix has more changes overall, but also causes additional bursts of changes as it moves around the group.

3.5 Conclusion

The three papers presented here have identified important affordances and tradeoffs based on interaction style and neighborhood definition, which we summarize here. First, using a



(a) Predator control

(b) Leader control

Figure 3.13: Power spectral density (PSD) of the B matrix. Leader-based control sustains influence better than predator-based control, as shown by the lower amount of power in the B changes across all frequencies.

topological rather than metric neighborhood decreased the amount of scrambling within the group. Second, parameter-based control performs well and requires little management, but must be optimized for the scenario *a priori*. Third, leader-based control sustains influence over the group, but the sustained influence causes more scrambling within the group than predator-based control. Fourth, predator-based control provided expressive behaviors such as the splitting of the group and interacted with fewer members of the swarm for a shorter duration per agent. However, predator-based control does not sustain influence and the expressive behaviors it affords require additional attention on the part of the human operator. Ultimately, the best choice of control method may depend on considerations for the scenario or task. The small control effort required by parameter-based control may be desirable for controlling massive numbers of agents, while the higher control effort required by the predator-based control may help the operator stay focused and vigilant during long-duration missions. Additional research is needed to better understand and quantify control method affordances and tradeoffs illustrated in this chapter and is further explored in Chapter 5.

The results presented in this chapter contribute foundational knowledge for continued research into the control and design primitives that form swarm systems. Chapter 4 and 5 continue to build on this work and expand our investigation to the impact on the human operator and individual swarm agents.

Chapter 4

Human-Swarm Interaction: User Study Design

4.1 Introduction

This chapter describes the goals, objectives, and implementation of the main user study presented in this thesis. The study is designed to directly evaluate the impact of HSI on the human-swarm system, including (a) individual agents, (b) the group as a whole, and (c) the workload of the human operator, as outlined in Section 1.2. Specifically, we use the data gathered from the user study to evaluate claims (3) and (4) which state that operator workload is affected by control style, but not the size of the swarm. The study also provides additional evidence for claim (2) regarding the effect of HSI on swarm topology. We evaluate these claims by designing a suitable information foraging task and monitoring both the swarm and human operator using a variety of metrics presented in this chapter.

The chapter is organized as follows: Section 4.2 discusses the high-level design of the study, including independent and dependent variables, the study scenario, scoring, metrics, and statistical considerations in the study design. Section 4.3 describes the implementation of the study, including modeling, model parameters, data gathering, user interfaces, detailed metrics, and survey questions. Section 4.4 describes the execution of the user study, including pilot studies, participant recruitment, training, and IRB approval. Chapter 5 provides an extensive analysis and discussion of the user study results.

4.2 Experiment Design

This section discusses the high-level design of the experiment abstracted from implementation details, which are discussed in Section 4.3. Section 4.2.1 gives an overview of the objective and independent variables examined in the experiment. Section 4.2.2 discusses dependent variables of interest. Section 4.2.3 discusses the swarm guiding task used in the user study. Section 4.2.5 and 4.2.4 detail the information gathering task and scoring for the scenario. Section 4.2.6 discusses the primary metrics for measuring workload. Section 4.2.7 examines statistical considerations in the design of the user study.

4.2.1 Objective and Independent Variables

The objective of this user study is to measure the impact of (1) interaction style and (2) scale on the mental workload of the human operator while controlling a swarm. Additionally, we also measure the impact of human interaction on the topology of the swarm and the trajectory stability of swarm members to reinforce the findings in Chapters 2 and 3. The experiment format is a 3 x 3 within-subject experiment where each participant performs all nine experimental conditions. Performance is compared between conditions for each participant and averaged across participants. We use three simple interactions styles – leader, predator, and stakeholders – as outlined in the section 1.4.5. For scale, we use 20, 50, and 100 simulated robots, which represents the range of swarm sizes commonly found in the literature.

We expect that swarms will scale well in terms of operator workload. This means that the workload of the human operator will remain relatively constant as more robots are added to the simulated swarm across all control styles. This is in contrast to traditional human-supervisory control, where the operator's workload increases with the number of robots [57]. We also expect that interaction style will affect both the performance of the swarm and the workload of the human operator. We expect that the performance of the swarm will increase as more robots are added, and that the additional robots will not cause the topology of the swarm to change more rapidly over time. In other words, we expect that increasing the size of the group will not cause the group to scramble.

4.2.2 Dependent Variables of Interest

We are primarily interested in the performance of the swarm on the study task and the mental workload of the study participants. We are also interested in the changes in the topology of the swarm over time. Swarm performance is measured by looking (a) at the amount of information gathered on the primary task and (b) at the user's score for the scenario. The participant's score on the scenario is calculated using the amount of information gathered and their performance on the visual secondary task. The change in topology of the swarm over time will be analyzed using power spectral density plots as explained in Chapter 3.

4.2.3 Study Scenario

In each condition of the experiment, the study participant is responsible for guiding simulated robots to points of interest around the screen. This represents a canonical information foraging task as outlined in Section 1.3.1. The number of robots and the control method are determined by the experimental condition. The points of interest are represented graphically on the screen by barrels and represent a task, such as surveillance, that can be completed in a finite amount of time and parallelized between robots. Two points of interest are initially present and additional points are generated randomly as the scenario progresses. The coordinates of each point of interest are also randomized.

The participant is also given two secondary tasks for the purpose of measuring workload, a visual response-time task, and an auditory memory and counting task. The secondary tasks were selected to measure the visual focus, mental processing, and response planning, of the study participant during the scenario. A score, representing the amount of information gathered and accuracy responding to the visual secondary task, is also shown on the screen. This helps the participant gauge his or her effectiveness on the task and maintain interest

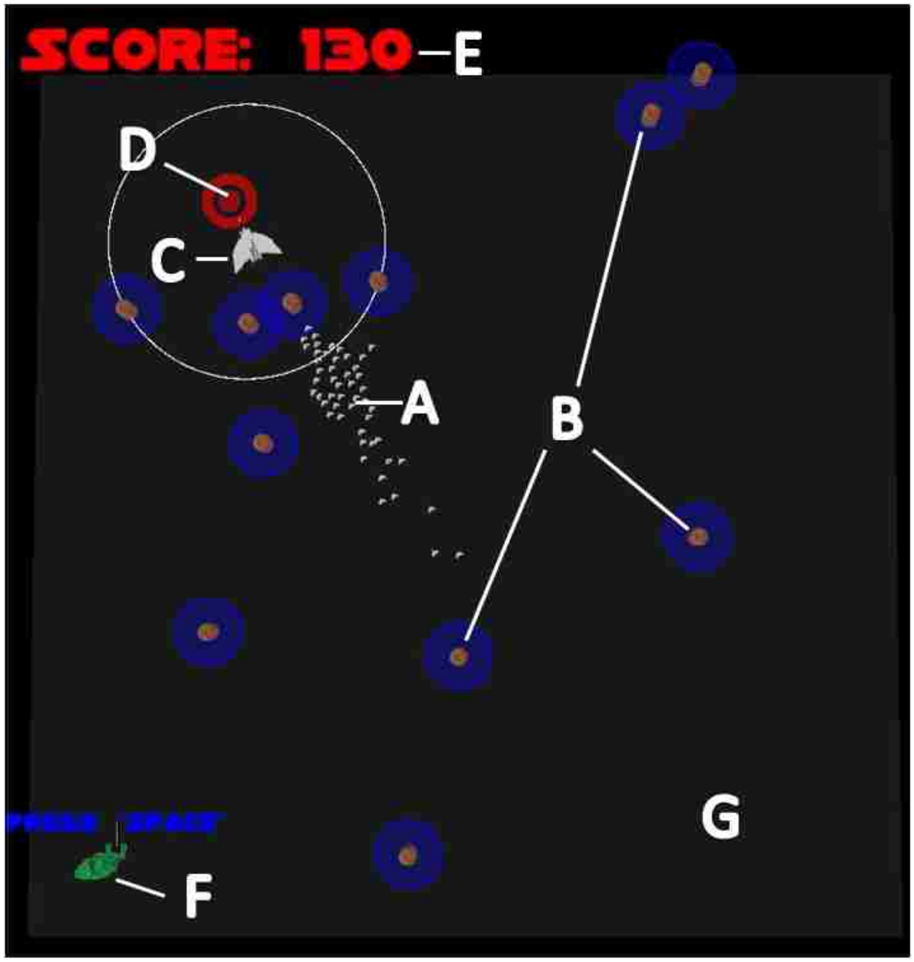


Figure 4.1: The information foraging task for the user study. The study participant is responsible for guiding a group of simulated robots (a) to areas of interest represented graphically by barrels (b), using leader, predator or stakeholder control (c). The leader, predator or stakeholders move towards the participant's cursor (d). The visual secondary task (f) is represented by a ninja that appears randomly on the playing field (g).

and focus during the experiment. Figure 4.1 shows an annotated screenshot of the scenario for the user study.

4.2.4 Score

The participant's score for the scenario is shown in the upper left hand corner of the graphical user interface and represents the participant's performance on that study scenario. The score starts at zero when the scenario begins and points are added or subtracted as the scenario progresses according to the following equation:

$$score = v_{recognized} - 5 * v_{missed} + I_g + T_{depleted} \tag{4.1}$$

where $v_{recognized}$ and v_{missed} are detections and missed detections on the visual secondary task, I_g is the total information gathered, and $T_{depleted}$ is the number of targets depleted.

The score primarily reflects the amount of information gathered, with small bonuses and penalties for secondary task performance. One point is added for each unit of information gathered by the simulated robots. A response to the visual secondary task event before the next event occurs is also worth one point, while failing to respond before the next event loses five points. This helps keep the participant focused on the study scenario and gives him or her feedback about his or her performance. The participant's accuracy on the auditory secondary task is calculated when the participant enters his or her guess after the end of the scenario. Consequently, the score, which is only shown while the scenario is running, does not include the auditory secondary task.

4.2.5 Primary Task

The study scenario uses a canonical information foraging task where the simulated robots move around the screen spatially and gather information at specific points of interest (targets). The amount of information gathered is proportional to the number of robots within range of the target and represents a variety of tasks that benefit from multiple robots, such as surveillance or search and rescue. The more robots that are within range, the faster the target is depleted. When all information has been gathered from the target, it disappears. The simulated robots do not seek targets on their own, but depend on the human operator to guide them to the targets.

The points are interest are represented graphically on the screen by barrels and are created at random locations (x, y) on the screen where x and y are independent and are distributed uniformly over the playing field. Initially, two points of interest are present on the screen and new points of interest spawn randomly during the scenario. Each time a new target is created, the amount of time (in seconds) until the next target appears τ_P is sampled from a uniform distribution over $[1.0, 11.0]$.

Each point of interest P_j contains an initial amount of information P_j^0

$$P_j^0 \sim u(10, 70) \tag{4.2}$$

and is depleted at a rate γ for each agent i within range R_p of the target

$$N_{P_j} = \{i : \|c_j - c_i\| \leq R_p\} \tag{4.3}$$

$$P_j(t+\tau) = \begin{cases} P_j(t) - \gamma N_{P_j}\tau & \text{if } P_j(t) \geq \gamma N_{P_j}\tau \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

The total amount of information gathered I_g is simply the amount of information depleted from all barrels during the simulation.

$$\Delta P_j = P_j(t) - P_j(t+\tau) \tag{4.5}$$

$$I_g(t+\tau) = I_g(t) + \sum_{j=1\ldots N_{points}} \Delta P_j \tag{4.6}$$

Table 4.1 lists the parameter values used in the user study.

Parameter	Description	Value
R_p	Information gathering radius	6.0 (24 meters)
γ	Information gathering rate	1.0
P^0	Initial information per target	$\sim u(10, 70)$
τ_P	Time until new target	$\sim u(1, 11)$

Table 4.1: Parameters for the primary information foraging task.

4.2.6 Primary Workload Measures

Secondary Task Design

According to Wickens' multiple resource theory [69], humans do not have a single information processing resource but have multiple specialized pools of information processing capabilities. These distinct pools represent sensory channels or other cognitive capabilities. Consequently, people will perform better on simultaneous tasks that require different channels than tasks that compete for the same channel. Thus it is important to consider different aspects of mental workload in the design of secondary tasks.

The primary task of controlling the swarm has two main components, (1) visual information processing and focus and (2) cognitive resources needed for decision-making and planning. The two secondary tasks are designed to monitor the demands placed on these mental resources. The first is a visual response-time task that requires the participant to press a key whenever a small ninja is displayed on the screen. This task competes for the participant's visual processing resources and visual attention but requires few cognitive resources. The second is an auditory counting and memory task that requires the participant to count the occurrence of one of two sounds. This task requires auditory resources to recognize the sound and cognitive resources to remember and count the number of occurrences of the sound. Because it uses sound rather than sight, it does not compete against the other secondary task, but does compete against the primary task for the participant's cognitive resources.

When properly calibrated, the two tasks combined provide objective information about the mental workload demands imposed by the task. If the demands on the visual or cognitive resources of the participant increase, there should be a corresponding decrease in performance on the secondary task as there are fewer resources available to perform it.

Auditory Secondary Task

The auditory secondary task uses two sounds, a cow sound and a robot sound. The participant is instructed to keep track of the number of times they hear the cow sound[1]. At the end of each scenario, they report their count and confidence in that count using a survey form. No sounds are played during the last 3 seconds of the scenario to avoid a partial sound being played.

The timing of the sounds and ratio of cow sounds to robot sounds is randomized. Each time a sound is played, the time (in seconds) until the next sound is drawn from a uniform distribution, $\tau_A \sim u(3.0, 7.0)$. The probability of a cow sound playing (vs a robot sound), is a Bernoulli random variable with parameter ρ, where $\rho \sim u(0.55, 0.75)$ and is fixed at the start of each scenario. This is done to keep the workload caused by the secondary task relatively constant between scenarios while varying the varying the number of cow sounds. This avoids confounding the results by (a) having the secondary task overly influence the workload measure being monitored, or (b) introducing a learning effect by keeping the number of cow sounds too constant, allowing the participant to accurately guess the answer.

Visual Secondary Task

The visual secondary task is designed to monitor the workload imposed by the visual attention and processing components of the primary task. This is accomplished with a response-time visual recognition task.

During the scenario, a graphic of a ninja will appear periodically at random locations in the graphical user interface (Figure 4.2). The participant is instructed to press the space bar each time they see a ninja. The location of the ninja is distributed uniformly over the playing field, in the same manner as the information targets described in Section 4.2.3. The time between when each ninja appears and when the space bar is pressed is recorded, along with the number of times the ninja appeared, the number of times the participant responded

[1]The volume of the two sounds was equalized.

Figure 4.2: A graphic of a ninja is periodically displayed at random locations on the screen. The participant must respond to the ninja by pressing the space bar. The participant's response time, false positives, and missed detections are recorded.

to the ninja and the number of times the participant did not respond to the ninja before the next ninja appeared[2]. There is at most one ninja on the screen at any given time. The time between ninja appearances is drawn from a uniform random variable $\tau_V \sim u(1.0, 4.0)$ in the same manner as the auditory secondary task.

NASA-TLX

The NASA Task Load Index (TLX) [40] is a subjective workload assessment tool that has been used extensively in HRI research. In the user study, it is used to rate the participant's perceived load during each of the experimental conditions. It is presented to the participant after each experimental condition and once following the three training scenarios to establish a baseline.

4.2.7 Statistical Considerations

The design of the user study was reviewed by Dr. Dennis Eggett from the BYU statistical consulting center and carefully analyzed before beginning the experiment. Several possible statistical issues and confounding factors were identified and steps were taken to mitigate these issues in the final experiment. Three areas in particular were identified: (1) intra-metric correlations, (2) introduced effects in experimental conditions, and (3) learning effect. The auditory secondary task required additional considerations.

[2] In this case, the previous ninja disappeared when the new ninja appeared.

Intra-metric Correlations and Introduced Effects

Each randomized event uses an independent pseudo-random number generator, which was seeded individually and only used for that event. The seeds were selected randomly at the start of each scenario, so that the sequencing of scenario events was unique per participant per scenario. Random seeds were also logged so that event flows could be reproduced and audited if needed. In order to keep the workload due to secondary tasks consistent within each scenario and across scenarios, we chose to randomize the duration between events, instead of using a finite probability of an event per tick. This allowed us to enforce minimum and maximum separation between events and avoid framerate dependent effects. Consequently, the randomization of each event is not affected by the frame rate of the simulation, the experimental condition, or the sequencing any other events.

Learning Effect

The study was expected to have a learning effect – improvement in performance over the duration of the experiment – and care was taken to minimize this effect and avoid it as a confounding factor in the results. First, a training scenario was designed to ensure proficiency with the control methods before the participant started the main experiment. The training is standards-based, meaning that the participant reached a designated level of proficiency before starting the main experiment. Second, the order of the training scenarios and experimental conditions were independently randomized to average out any learning effects over the course of the experiment.

Auditory Secondary Task

Special care was taken with the auditory secondary task, in which the participant counts the number of times a cow sound has occurred during the scenario. The elapsed time between events is a uniform random variable, and the number of events is a function of the sum of the timings. By the central limit theorem, the distribution over the number of cow sounds during each scenario will approximate a Gaussian. Without any additional steps, the participant could have guessed the count with high accuracy instead of counting sounds as instructed. The reduced workload over time would have introduced a strong learning effect into the experiment.

To remedy this, we randomized the ratio of cow sounds to robot sounds between scenarios, while holding the distribution over time between sounds constant. The ratio of cow sounds to robot sounds is fixed at the start of each scenario and drawn from a uniform distribution over $[0.55, 0.75]$. Consequently, the workload due to the secondary task stays relatively constant per scenario (roughly the same number of audio events), but the number of cow sounds varies.

4.3 Experiment Implementation

This section discusses the implementation details of the user study including modeling, model parameters, user interface, and an overview of metrics and surveys used in the experiment. Section 4.3.1, 4.3.2, and 4.3.3 detail the model, model parameters, and control model used in the user study. Section 4.3.4 discusses the graphical user interface. Section 4.3.5 and 4.3.6 discuss metrics and surveys used in the user study. Section 4.3.7 discusses the data gathered during the user study. Section 4.3.8 discusses implementation details such as the programming language and libraries used.

4.3.1 Swarm Model

The user study uses the augmented version of Couzin's model outlined in section 1.4.6. This model is a canonical flocking model suitable for the user study task and provides a reasonable model for the coordinated movement of UAVs. It uses dynamics very similar to Dubins airplane [17], a simple plane model, and provides a minimum separation distance (zone of repulsion), similar to aircraft visual flight rules (see [70] for a discussion). While Couzin's model does not provide guarantees for collision avoidance, we assume that the UAVs can change altitude to avoid collisions. Call et al. [14] provide an example of UAV collision avoidance in the field using on-board sensors.

4.3.2 Selection of Model Parameters

The initial aim of the experiment was to simulate the TurtleBot robots that we have in our lab. However, simulations showed that the Turtlebots created wide, slow moving flocks unsuitable for short-duration experiments. This is due to their limited speed and localization accuracy relative to their size. Consequently, we instead model a flock of simulated UAVs in order to keep the simulation tied to a real-world system. However, we note that the model and study task are not UAV-specific and apply to a variety of swarm robot systems, including ground and underwater vehicles.

The parameters for the model were chosen to match the low-cost UAVs used by the MAGICC lab (Figure 4.3) during flight tests, including forward speed, turning rate, minimum recommended separation, object detection range using onboard cameras, and GPS localization error. UAVs match the dynamics of Couzin's model more closely and provide additional experimental validity. Table 4.2 contains the parameters values for the augmented version of Couzin's model found in Section 1.4.6. In the simulation, 1 unit is equal to 4 meters and the parameters from the UAV flight tests were scaled accordingly.

Modeling of Localization Error

GPS error is difficult to model and depends on satellite positioning, local weather, terrain features and a variety of other error sources [10]. The GPS error drifts over time as terrain, satellite position and other factors change. However, high accuracy measurements are possible between two relatively close GPS units [58]. Because of this, we use a simplified error model in the user study. The error model is general and can also be applied to visual navigation

[3] Photo courtesy of the BYU MAGICC lab.

[4] u-blox NEO-6p GPS receiver board with SMA. Photo courtesy of CSG Shop and used with permission. `http://www.csgshop.com/`

(a) A small UAV[3]

(b) A low-cost GPS unit[4]

Figure 4.3: Small UAVs, such as those used in the BYU MAGICC lab (a), are commonly equipped with GPS, onboard cameras, and processing hardware. The parameters for the user study match those used in similar UAVs during flight tests. Small low-cost GPS units, such as the unit shown in (b), are capable of 1m global accuracy. More expensive augmented dual-signal units are capable of 10cm accuracy.

Parameter	Description	Value	SI
s	Forward velocity	4.0 units/s	16.0 m/s
θ	Turning rate	45.0 degrees/s	45.0 degrees/s
α	Blind spot angle	45.0 degrees	45.0 degrees
R_r	Zone of repulsion radius	1.5 units	6 meters
R_o	Zone of orientation radius	11.0 units	44 meters
R_a	Zone of attraction radius	14.0 units	56 meters
w_g	Stakeholder control gain	20.0	-
w_p	Leader/predator control gain	25.0	-
R_p	Leader/predator interaction radius	25.0 units	100 meters

Table 4.2: Simulation parameters for the model UAVs. Parameters match those used in UAV flight tests by the BYU MAGICC lab (see Figure 4.3).

systems, such as that shown in [4]. Small, inexpensive GPS units with 1 meter global accuracy and suitable for use in UAVs are currently available (see [13, 41] and Figure 4.3b). More expensive units capable of global centimeter-level accuracy are also commercially available (see [29]).

Based on the discussion above and our tests from group robots in Section 2.5, we modeled the localization error of the simulated UAVs as a bivariate Gaussian distribution with a standard deviation of 1.0 meters (0.25 simulation units). This error is held constant for a period of τ_ϵ seconds drawn from a uniform distribution $\tau_\epsilon \sim u(1, 20)$. We believe that this model overestimates the relative GPS error between simulated UAVs.

4.3.3 Control Model and Parameters

Leader and Predator

The leader and predator represent points of influence under the control of a human operator. The leader and predator use identical control laws and move towards the participant's mouse using a PD controller with a maximum acceleration and velocity. Table 4.3 shows the controller values used in the user study.

Parameter	Description	Value
P	Error gain	1.0
D	Derivative gain	-2.0
a_{max}	Maximum acceleration	10.0
v_{max}	Maximum velocity	30.0

Table 4.3: Parameter values for the leader/predator PD controller.

Stakeholders

For stakeholder control, the location of the participant's mouse is passed as a goal location to the stakeholders as outlined in Equation 1.13, using the parameters in Table 4.2.

4.3.4 User Interface

This section describes the design of the graphical user interface for the user study. The goal of the interface is to (1) display information about the swarm, study scenario, and participant performance, and (2) allow the participant to influence the swarm.

Primary Display

The primary purpose of the graphical interface is to provide spatial information about the simulated robots and locations of interest (see Figure 4.4). The playing field or area where points of interest spawn is shown by a large gray rectangle. Points of interest are represented as barrels surrounded by a circle that represents the minimum distance for robots to gather information from that point. The robots themselves are displayed as small planes, while the leader and predator are displayed as a large plane or shark respectively. Agents under the participant's control are surrounded by a white circle that represents their radius of influence.

Control Input

The study participant controls a target-shaped cursor on the screen. The agents under the participant's control move toward this target as it is moved around the screen with the mouse.

Figure 4.5: The cursor

4.3.5 Metrics

Performance Metrics

The metrics in this category measure the performance of the swarm in each condition to measure the impact of scale and control style. The participant's score, total information gathered, and targets depleted are monitored.

Metric	Description	Units
Information gathered	Total information gathered during the scenario	agent-seconds
Targets depleted	Number of targets completely depleted	(count)
Score	The participant's score for the scenario	(none)

Table 4.4: Performance metrics

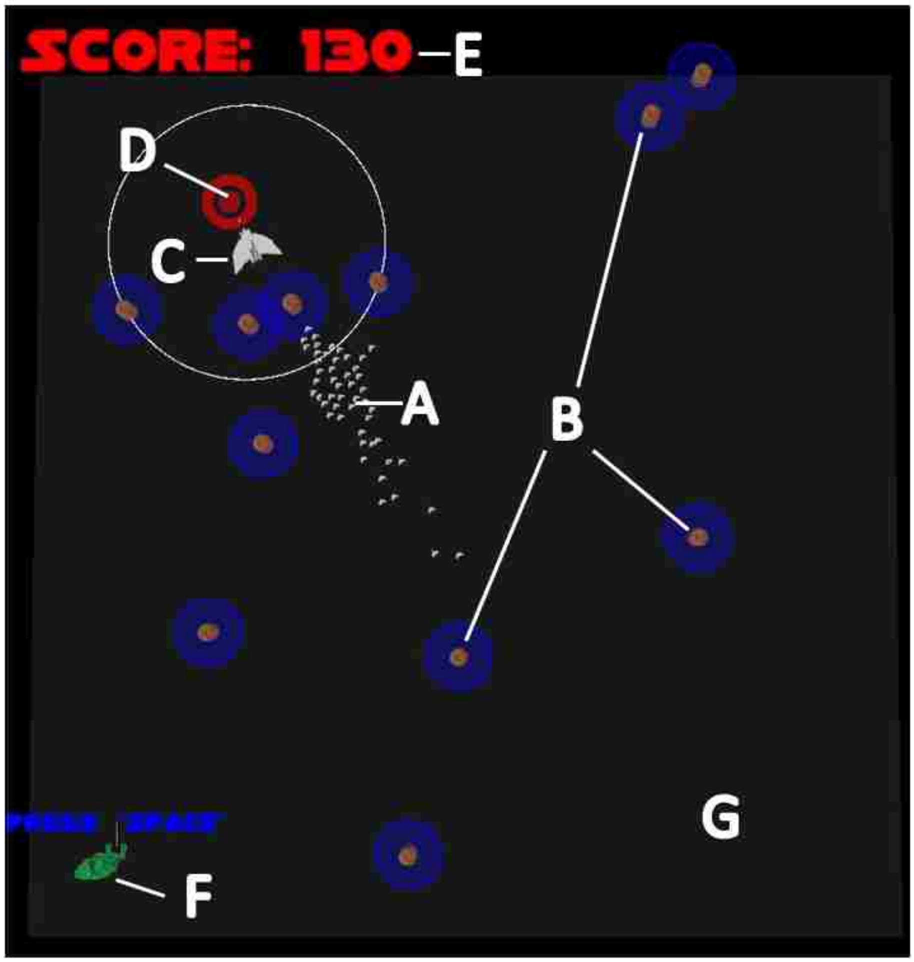


Figure 4.4: The information foraging task for the user study. The study participant is responsible for guiding a group of simulated robots (a) to areas of interest represented graphically by barrels (b), using leader, predator or stakeholder control (c). The leader, predator or stakeholders move towards the participant's cursor (d). The visual secondary task (f) is represented by a ninja that appears randomly on the playing field (g).

Objective Workload Metrics

The metrics in this category provide objective measurements of the participant's workload during each study scenario. They are divided into three categories: control activity, auditory secondary task, and visual secondary task. Control activity is measured by tracking the participant's mouse during the study scenario.

Metric	Description	Units
Total Mouse Movement	The total distance the mouse moved during the scenario	pixels

Table 4.5: Control activity metrics

Metric	Description	Units
True Positives	The number of cow sounds played during the scenario	(count)
Guess	The participant's estimate of the number of cow sounds	(count)
Guess Ratio	Ratio of the participant's estimate to the true value	(none)
Percent Error	Percent error of the participant's estimate	percent
Confidence	The participant's confidence in his or her estimate	(survey)

Table 4.6: Auditory secondary task metrics

Metric	Description	Units
False Positives	The number of false positives on the visual task	(count)
True Positives	The number of correct responses on the visual task	(count)
Response Time (no MD)	Visual task response time (excluding missed detections)	seconds
Response Time	Visual task response time (including missed detections)	seconds

Table 4.7: Visual secondary task metrics

NASA-TLX

Table 4.8 shows the six components of the NASA-TLX rating scale. Additionally, fifteen pairwise comparisons are used to weight the importance of each scale.

4.3.6 Survey Questions

Participants complete three surveys after each experimental condition. First, they report the number of cow sounds they counted and their confidence in their answer (Table C.3). Second,

Metric	Description	Units
Temporal Demand	Participant's rating for TLX temporal demand	20 point scale
Effort	Participant's rating for TLX effort	20 point scale
Physical Demand	Participant's rating for TLX physical demand	20 point scale
Frustration	Participant's rating for TLX frustration	20 point scale
Performance	Participant's rating for TLX performance	20 point scale
Mental Demand	Participant's rating for TLX mental demand	20 point scale
Overall	The overall TLX score	100 point scale

Table 4.8: NASA-TLX rating scale

they complete NASA-TLX (Table 4.8). Third, they complete a short workload survey about their experience during that scenario (Table C.3). Additionally, participants complete a pre-experiment demographic questionnaire (Table C.2) and a post-experiment preference survey (Table C.1). The demographics were analyzed for fixed effects (e.g. video games experience). The post-experiment survey results were tallied and read, but not analyzed statistically. A full listing of the survey questions can be found in Appendix C.

Neglect Time

An important concept in HRI is neglect time, or the time that the human operator does not spend actively managing the robot [56]. In this user study, the movement of the participant's mouse is analyzed. Any time step where the mouse does not move is considered neglect time. We measure the duration and number of neglect intervals.

Metric	Description	Units
Total Neglect Time	Sum of all time intervals with no mouse movement	seconds
Longest Neglect Interval	Longest time period with no mouse movement	seconds
Average Neglect Interval	Total neglect time divided by the number of neglect intervals	seconds
Number of Neglect Intervals	Number of time periods with no mouse movement	(count)

Table 4.9: Neglect time metrics

Possible Confounding Factors

As mentioned in Section 4.2.7, care was taken to avoid confounding factors in the design of the user study. Additionally, several of the metrics of concern were monitored and analyzed for fixed effects (see Table 4.10).

Metric	Description
Experiment ordering	Learning effect during the experiment
Framerate hiccups	Number of simulation ticks longer than 0.5 seconds
Number of visual events	Number of times the ninja appeared
Number of audio events	Total number of sounds played
Audio ρ	Ratio of robot sounds to cow sounds
Audio true positives	Number of cow sounds played
Audio false positives	Number of robot sounds played

Table 4.10: Possible confounding factors

4.3.7 Data gathering

All data collected was anonymized and stored on disk using a per-participant unique identifier generated at the start of the experiment. The participant's name appears only on the authorization form, which was stored separately and not associated with the identifier. Data gathered during the experiment included survey responses, the ordering of experimental conditions, and interaction during each experimental condition. All data was logged to text files and then archived. Scripts were used to parse the raw data and provide experimental results.

All data recorded was marked with two timestamps, the elapsed wall-clock time and simulation time since the start of the scenario. All significant events during the scenario were recorded including dialog interaction, pauses, key presses, mouse clicks, changes in framerate, and events related to the secondary tasks. Additionally, the location, heading, desired heading and localization error of each robot, leader or predator where recorded at 0.10 second intervals. The location of the mouse and total distance moved so far was recorded at the same rate.

4.3.8 Implementation Details

The implementation of the simulator was identical to the description in Chapter 3 with additional modifications for dialogs, user interaction and stakeholder control. The training video was created with Microsoft PowerPoint 2010 using screenshots from the interface and freely available graphics. All of the surveys were coded in PHP and rendered via the Firefox web browser. The automation for the user study was done via a python script that launched the simulations and surveys. Analysis of the data was done using Perl, Matlab 2010a, SAS 9 and Microsoft Excel 2010. All of the plots were generated with Matlab. The simulation for the user study was coded in C++ using the open-source OGRE graphics library and CEGUI user interface library.

4.4 Experiment Execution

This section describes the execution of the experiment including pilot studies (Section 4.4.1), participant recruitment (Section 4.4.2), and the flow of the experiment (Section 4.4.3).

4.4.1 Pilot

A short pilot study was completed prior to the main user study (1) to expose any bugs in the computer programs used in the user study, (2) to expose any confounding factors in the study and (3) to verify that the expected trends were present in the data.

Four participants were recruited from my research lab to pilot the user study. All of them completed the study successfully with no reported side effects. The data was analyzed by a statistical consultant, Dr. Dennis Eggett, and the expected trends were present. As a result of feedback during the pilot, several dialogs were added or clarified and the training was updated.

4.4.2 Participant Recruitment

Participants for the user study were recruited via flyers, email and word of mouth. Flyers were posted to the stairwells in various buildings at Brigham Young University. An email containing the flyer was sent to (1) the mailing lists for undergraduate classes taught by Dr. Michael Goodrich (my advisor) (2) employees of the Humanities Learning Resources lab supervisor and (3) members of my local church unit. Participants were also recruited by word of mouth in my church unit and the BYU MAGICC lab. A copy of the flyer is shown in Appendix B. None of the participants had knowledge of the hypothesis or goals of the user study prior to the experiment.

4.4.3 Flow of the Experiment

Summary

The experiment took approximately one hour per participant and took place in the HCMI lab. Each participant viewed a short training video with instructions, completed an initial demographics questionnaire, and then a training with each control method. After the training, they completed all of the nine experimental conditions in groups of three. The order of the conditions was randomized, as was the order of the training scenarios. Between each group of scenarios, the participant was given a 5 minute break. At the end of the experiment, each participant completed a post-experiment survey and was thanked for his or her participation.

Location and Equipment

The user study was conducted in the HCMI lab of the Computer Science department at Brigham Young University. Two participants from the BYU MAGICC lab ran the study in their own lab using the laptop configured for the study.

Two computer were used for the user study, one desktop and one laptop computer which ran the software using VirtualBox virtual machine software. Both computers were running Ubuntu Linux 12.04 and used identical library and software versions.

Preliminary

The user study was expected to take approximately one hour. Almost all of the study participants were able to complete it in one hour and those that didn't only went a few minutes over.

When each participant arrived, they were greeted and given the IRB consent to be a research subject form and the experiment payment form. Each participant was offered compensation of $12 for his or her participation. If the participant wished to decline payment, he or she was given an alternate IRB form, which specified no additional compensation.

Training and Practice

After completing the forms, the participant was shown a five minute training video about the goal of the scenario (obtain as many points as possible), the user interface, control methods, and secondary tasks.

After the training video, the participant completed a short demographics questionnaire and began the experiment. They were first given a chance to practice with each control method and complete a practice NASA-TLX questionnaire. Each training scenario has two parts, (1) guide the simulated robots to each of four barrels spread over and playing field and (2) get a score of 100 points by guiding the robots to barrels and responding to the secondary tasks. This ensured that each participant had a minimum level of proficiency guiding the robots and responding to the secondary tasks.

Experimental Conditions

The order of the nine experimental conditions are randomized and then broken into groups of three. The participant completes the nine conditions with a five minute break between groups. Participants were allowed to take more or less than five minutes if they wanted. After each scenario, each participant (1) was asked the number of cows sounds and confidence, (2) completed NASA-TLX, and (3) completed a four question post-scenario survey. The participant was allowed to skip any questions he or she did not feel comfortable answering.

Each experiment scenario lasts for two minutes of simulation time, not including time spent in dialogs or paused. When the scenario starts, a prompt with instructions is displayed. When the dialog is dismissed, the scenario begins after five seconds. The delay provides adequate time to move the mouse cursor to a desirable location. At the end of the scenario, the user interface closes. The surveys are displayed one at a time in order.

Post-Experiment

After completing all nine scenarios, each participant completes a post-experiment survey where they are asked about their preferred control method and number of agents and invited to comment.

4.5 Summary

This chapter provides a detailed description of the design, implementation, and execution of the user study along with the experimental, modeling, and statistical considerations taken into account. The study was carefully designed based on our prior research presented in Chapter 3 and takes steps to avoid confounding factors and ensure measurable results. The results of the study are presented in Chapter 5.

Chapter 5

Human-Swarm Interaction: User Study Results and Discussion

5.1 Introduction

In this chapter we present, analyze, and discuss the results from the user study described in Chapter 4. Section 5.2 describes the data gathered in the user study, the datasets not used, and reported side effects. Section 5.3 reports the statistical analysis of the results of the user study, organized by group of metrics, along with an analysis of possible confounding factors, learning effects, and demographics. Section 5.4 synthesizes the results from the statistical analysis and presents overall findings along with implications for future research and design.

5.2 Data Gathered

In this section, we discuss the results from the user study, including reported side effects, the data that was gathered, and the datasets we used.

5.2.1 Reported Side Effects

There were minimal side effects reported during the user study. One participant reported watery eyes from concentrating on the computer screen and discontinued the experiment in accordance with the IRB protocol. The symptoms were minor and abated once the experiment was discontinued. No other side effects were reported.

5.2.2 Excluded Data

We collected data from all 32 volunteers that participated in the user study. We chose to exclude 5 datasets from our analysis. One dataset was incomplete due to study side effects (see Section 5.2.1). Three datasets were excluded due to computer problems that resulted in incomplete data, or required that participants move to another computer. One additional dataset was not used due to deviation from the instructions for the user study. The analysis of the remaining 27 datasets is presented in Section 5.3.

5.3 Analysis

In this section, we conduct a statistical analysis of the user study data and identify significant results. We analyze possible confounding factors, learning effects, and demographics and present analysis for each group of metrics, as arranged in Section 4.3.5. Statistical data is presented in this section as needed. The data from the complete statistical analysis, including fixed effects, means, and pairwise tests, is found in Appendix D.

5.3.1 Statistical Analysis

The data was compiled and analyzed by Dr. Dennis L. Eggett, a statistical consultant in the BYU Department of Statistics, using SAS 9. The statistical analysis was conducted using mixed models analysis with Tukey-Cramer adjustment. We analyzed pairwise within-subject differences across control style and scale. We also used mixed model analysis to check for fixed effects across scale, control style, demographics, scenario ordering (learning effect), and several confounding factors. Table 5.1 lists the fixed effects we analyzed and references the relevant sections of this thesis.

All tables in this section are in the format *significance* or *mean* ± *standard_deviation (significance)*, where *mean* represents the mean or difference in means and *significance* is the statistical significance of the Tukey-Kramer adjusted p-value. In this thesis we use a

Effect	Type	Section
Scale	Independent variable	4.2.1
Style	Independent variable	4.2.1
Sex	Demographic	4.3.6
Vision	Demographic	4.3.6
Robot Exp	Demographic	4.3.6
Games Exp	Demographic	4.3.6
Sim Order	Learning effect	4.2.7

Table 5.1: Fixed effect tests

significance value of $p < .01$, while making note of effects where $p < .05$, which we refer to as weakly significant.

5.3.2 Possible Confounding Factors

As described in Section 4.2.7, a number of possible confounding factors were identified in the initial design of the user study and steps were taken to mitigate them. These possible confounding factors were analyzed after the completion of the user study to check for any effects, as shown in Table 5.2. Only one factor, framerate hiccups – the number of significant drops in framerate during the simulation – was significant for either independent variable. This effect was anticipated, as the amount of computation and data logging increases with the number of simulated robots.

We further analyzed framerate hiccups with pairwise comparisons between conditions (as shown in Table 5.3) and a review of other metrics to check for any negative effects. It was expected that a substantial number of framerate hiccups would degrade performance or increase workload. However, we found that (1) the number of hiccups was very small (1-2 events per scenario), and (2) none of the expected negative effects were present in the data. For all other metrics, scale was either not significant, or the metric improved with scale. Consequently, we conclude that none of the possible confounding factors we identified,

Metric	Scale	Style
Number of Visual Events	0.4677	0.3557
Experiment Ordering	0.5456	0.0759
Framerate Hiccups	$< \mathbf{.0001}$	0.4656
Number of Audio Events	0.3390	0.9972
Audio p(False Alarm)	0.1986	0.8837
Audio True Positives	0.2897	0.7067
Audio False Positives	0.4477	0.7772

Table 5.2: $Pr > F$ for fixed effects tests on possible confounding factors. Of the confounding factors we identified, only framerate hiccups was significant. We conducted additional analysis (Table 5.3) and concluded that the results of the user study were not confounded by framerate hiccups or any other factor we identified. See Table 4.10 for a description of the entries in this table. Bold text indicates statistical significance.

	Scale		
Metric	20 vs 50	20 vs 100	50 vs 100
Framerate Hiccups	$\mathbf{-1.50 \pm 0.23}$ ($< \mathbf{.0001}$)	$\mathbf{-1.94 \pm 0.23}$ ($< \mathbf{.0001}$)	-0.44 ± 0.23 (0.1390)

Table 5.3: Pairwise analysis of possible confounding factors across scale. We further analyzed framerate hiccups to determine if it had confounded the user study results and found that (1) the difference between conditions is small (1-2 events per scenario) and (2) none of the expected negative effects are present in other metrics. We therefore conclude that the results of the study were not confounded by framerate hiccups or any other factor we identified.

including framerate hiccups, had a significant impact on the results of the user study and that no additional adjustments are required for the remainder of the analysis.

5.3.3 Learning Effect

As discussed in Section 4.2.7, we expected that some of the metrics used in the user study would show a learning effect – improvement in performance over time – and took steps to minimize this effect. We analyzed the data after the conclusion of the experiment to (1) check for learning effects and (2) verify that any learning effects identified did not confound or bias the results of the study.

Metric	Scale	Style	Sim Order
Experiment Ordering	0.5456	0.0759	
Visual Missed Detections			**0.0022**
Visual True Positives			**0.0029**
Audio Guess Error (absolute)			**0.0347**
Number of Neglect Intervals			**0.0110**

Table 5.4: Learning effect and correlation with study independent variables. This table shows the statistical significance of simulation order for both independent variables and $Pr > F$ for fixed effect tests for metrics that show a learning effect. We found no correlation between simulation order and our independent variables, and identified a single statistically significant ($p < .01$) learning effect on the visual secondary task. Bold text indicates statistical significance.

As shown in Table 5.4, the number of missed detections on the visual secondary task had a statistically significant learning effect[1]. Interestingly, this was a slight decrease, rather than increase, in performance over the course of the experiment. Based on participant comments, we hypothesize that participants became more focused on the primary task to attempt to obtain a higher score as the experiment progressed. We note however, that this difference was small, at approximately 1-2 missed events per scenario and did not have a significant impact overall.

We also found several other weak learning effects ($p < 0.05$) in the data. Participants showed a slight improvement on the auditory secondary task and a slightly smaller number of neglect intervals per scenario over the duration of the experiment. This may be indicative of slight adjustments participants made to their control strategy over time, as suggested by participant comments.

Most importantly, we found that simulation order was not correlated with (1) either independent variable, (2) performance, or (3) surveys, including NASA TLX. This means (1) the randomization of scenarios was sufficient to prevent learning effects from confounding the results, (2) the level of proficiency provided by the standard-based training was sufficient to avoid learning effects on participant performance, and (3) participants answered the surveys

[1]The complimentary statistic, correct visual secondary task responses, also had a statistically significant learning effect.

Metric	Sex	Vision	Robot Exp	Games Exp
Information Gathered	0.6078	**0.0085**	**0.0356**	**0.0212**
Audio Guess Error	**0.0118**	0.5916	**0.0279**	0.6197
Audio Guess Ratio	**0.0113**	0.2999	**0.0094**	0.6636
Number of Neglect Intervals	**0.0147**	0.6164	0.4226	0.2911
TLX Overall	0.0714	0.9030	0.4216	**0.0339**

Table 5.5: $Pr > F$ for fixed effect tests demographic questions. We found that (1) participants with corrected-to-normal vision gathered more information per scenario, and (2) participants with more robot experience scored slightly better on the auditory secondary task. Several other weak effects were also identified. Bold text indicates statistical significance.

consistently over time. The results from this section and Section 5.3.2 show that the design of the user study was sound and was not confounded by any learning effects or design issues identified in the initial plan for the user study (Section 4.2.7).

5.3.4 Demographics

We analyzed the results from the demographic survey described in Table C.2 and found two statistically significant effects ($p < 0.01$), as shown in Table 5.5. First, participants who reported corrected-to-normal vision gathered more information per scenario than those who did not. Second, those who reported more robot experience had a smaller error ratio on the auditory secondary task.

We also identified several weak effects ($p < 0.05$). Participants who reported high levels of robot experience or video games experience tended to gather more information during study scenarios. Additionally, video games experience was also correlated with lower reported workload on the NASA-TLX survey. A difference in the number of neglect intervals, but not total neglect time, was observed across sex, which suggests a slight difference in control strategies. A difference across sex was also observed in the audio guess ratio.

5.3.5 Swarm Performance

Both scale and control style significantly affected swarm performance on all measures ($p < 0.001$, see Table 5.6). The size of the effect was more pronounced for scale ($15 - 20\%$) than

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Information Gathered	< **.0001**	**0.0002**	0.6078	**0.0085**	**0.0356**	**0.0212**	0.4298
Barrels Depleted	< **.0001**	< **.0001**	0.4840	0.0583	0.0937	0.0829	0.6665
Score	< **.0001**	< **.0001**	0.5140	0.1306	0.1923	0.0944	0.5211

Table 5.6: $Pr > F$ for fixed effect tests on the performance metrics used in the user study. We found that scale and control style both significantly affected performance ($p < .0001$) across all of our metrics. We also found that corrected-to-normal vision, robot experience, and video games experience were correlated with higher performance. See Table 4.4 for a description of the entries in this table. Bold text indicates statistical significance.

for control style (10%), but both effect sizes are large enough to be of practical significance. Corrected-to-normal vision, along with higher robot and video games experience were also correlated with increased performance. A learning effect was not found on any of the performance metrics we measured.

Scale

Larger groups sizes increased swarm performance, but provided diminishing returns as the size of the swarm increased. As Table 5.7 shows, the performance difference between 20 and 50 agents was significant on all measures ($p < .0001$), but the difference between 50 agents and 100 agents was not significant on any measure. Additionally, increasing the group size from 20 to 50 shows a substantial increase in performance, but further increasing the group size to 100 provides a much smaller benefit, even though a larger number of agents have been added. As with many real-world systems, the information foraging task in the user study shows performance gains as more agents are added, but diminishing returns as the number of agents approaches a saturation point. The performance across scale matched our hypothesis.

Control Style

Swarm performance was also significantly influenced by control style. Leader-based control performed best on all performance measures, followed by stakeholder-based control, and then predator-based control. Leader-based and stakeholder-based control performed similarly and

Metric	Scale 20 vs 50	20 vs 100	50 vs 100
Information Gathered	$\mathbf{-72.86 \pm 12.07}$	$\mathbf{-96.47 \pm 12.00}$	-23.61 ± 12.13
	($\mathbf{< .0001}$)	($\mathbf{< .0001}$)	(0.1366)
Barrels Depleted	$\mathbf{-2.62 \pm 0.29}$	$\mathbf{-3.05 \pm 0.29}$	-0.43 ± 0.29
	($\mathbf{< .0001}$)	($\mathbf{< .0001}$)	(0.3209)
Score	$\mathbf{-80.46 \pm 12.16}$	$\mathbf{-101.74 \pm 12.08}$	-21.27 ± 12.22
	($\mathbf{< .0001}$)	($\mathbf{< .0001}$)	(0.2003)

Table 5.7: Pairwise comparisons of performance metrics across scale. Bold text indicates statistical significance.

Metric	Control Style L vs P	L vs S	P vs S
Information Gathered	$\mathbf{55.19 \pm 12.12}$	28.80 ± 12.22	-26.39 ± 12.04
	($\mathbf{< .0001}$)	(0.0572)	(0.0826)
Barrels Depleted	$\mathbf{2.46 \pm 0.30}$	0.40 ± 0.30	$\mathbf{-2.06 \pm 0.30}$
	($\mathbf{< .0001}$)	(0.3926)	($\mathbf{< .0001}$)
Score	$\mathbf{75.67 \pm 13.05}$	24.65 ± 13.13	$\mathbf{-51.02 \pm 12.98}$
	($\mathbf{< .0001}$)	(0.1561)	(**0.0007**)

Table 5.8: Pairwise comparisons of performance metrics across control style. Bold text indicates statistical significance.

did not show statistically significant differences. Differences between predator-based and leader-based control were always significant, while differences between predator-based and stakeholder-based control were significant for score and barrels gathered but not information gathered. These results show that design choices, such as control method, can influence the performance of the swarm and should be taken into consideration in the design of swarm systems.

5.3.6 Objective Workload Measures

In this section we analyze participant performance on the secondary tasks, and control activity, as measured by the total amount of mouse movement during the scenario. Scale did not have a statistically significant effect on either secondary task or on the total mouse movement. Interaction style significantly affected total mouse movement, performance on the visual secondary task, and the participant's confidence rating on the auditory secondary

Metric	Scale	Style	Sex	Vision	Robot	Games	Order
Audio Guess	0.2783	0.5195	0.4895	0.0924	0.3137	0.4057	0.4213
Audio Guess Error	0.9836	0.5891	**0.0118**	0.5916	**0.0279**	0.6197	0.0971
Audio Guess Error (absolute)	0.3008	**0.0336**	0.2572	0.8301	0.7109	0.7730	**0.0347**
Audio Guess Ratio	0.9464	0.5673	**0.0113**	0.2999	**0.0094**	0.6636	0.0547
Audio Percent Error	0.1784	**0.0331**	0.3685	0.6043	0.5806	0.6325	**0.0355**
Audio Confidence	0.4684	**0.0009**	0.6270	0.6170	0.3272	0.5046	0.0544
Visual False Positives	0.3192	0.5173	0.1158	0.5382	0.4833	0.0971	0.1538
Visual True Positives	0.4475	$<$ **.0001**	0.6614	0.6197	0.7104	0.3518	**0.0029**
Visual Missed Detections	0.2338	**0.0001**	0.6664	0.6371	0.5867	0.3581	**0.0022**
Visual Response Time (no MD)	0.0821	$<$ **.0001**	0.7379	0.6105	0.6596	0.5936	0.0739
Visual Response Time	0.1778	$<$ **.0001**	0.9092	0.5976	0.5745	0.4525	0.0515
Total Mouse Movement	0.3934	$<$ **.0001**	0.3555	0.0796	0.1836	0.9317	0.4031

Table 5.9: $Pr > F$ for fixed effects tests on objective workload metrics. Control style significant affected control activity, performance on the visual secondary task, and confidence on the auditory secondary task. Scale was not statistically significant on any measure. See Tables 4.7, 4.6, and 4.5 for a description of the entries in this table. Bold text indicates statistical significance.

task ($p < 0.001$, see Table 5.9). Robot experience and sex had strong and weak significance respectively on the audio guess ratio, but not overall error, as described in Section 5.3.4. Additionally, a significant learning effect was observed for missed detections on the visual secondary task (see Section 5.3.3). The remainder of this section provides a more detailed analysis of each group of metrics.

Auditory Secondary Task

Control style and scale did not significantly affect performance on the auditory secondary task due to a small effect size and high variance. However, participant's rating of their confidence on the secondary task was significant for interaction style ($p < .001$, see Table 5.10). Participant confidence for leader-based and stakeholder-based control was nearly identical ($\Delta\mu = 0.03$, $p = 0.95$), and was higher than predator-based control ($p < .005$). These results indicate that while accuracy on the secondary task did not change, predator-based control likely consumed more cognitive resources, resulting in decreased confidence on

Metric	Control Style L vs P	L vs S	P vs S
Audio Confidence	$\mathbf{0.41 \pm 0.12}$ ($\mathbf{0.0021}$)	0.03 ± 0.12 (0.9535)	$\mathbf{-0.38 \pm 0.11}$ ($\mathbf{0.0048}$)

Table 5.10: Pairwise comparisons across control style for the auditory secondary task. Bold text indicates statistical significance.

Metric	Control Style L vs P	L vs S	P vs S
Visual True Positives	$\mathbf{3.00 \pm 0.88}$ ($\mathbf{0.0038}$)	-1.16 ± 0.88 (0.3973)	$\mathbf{-4.16 \pm 0.88}$ ($< \mathbf{.0001}$)
Visual Missed Detections	$\mathbf{-2.92 \pm 0.81}$ ($\mathbf{0.0020}$)	0.55 ± 0.81 (0.7749)	$\mathbf{3.48 \pm 0.81}$ ($\mathbf{0.0002}$)
Visual Response Time (no MD)	$\mathbf{-0.07 \pm 0.02}$ ($\mathbf{0.0014}$)	0.03 ± 0.02 (0.3075)	$\mathbf{0.10 \pm 0.02}$ ($< \mathbf{.0001}$)
Visual Response Time	$\mathbf{-0.15 \pm 0.04}$ ($\mathbf{0.0003}$)	0.05 ± 0.04 (0.3554)	$\mathbf{0.20 \pm 0.04}$ ($< \mathbf{.0001}$)

Table 5.11: Pairwise comparisons across control style for the visual secondary task. Bold text indicates statistical significance.

the competing secondary task. This idea is further reinforced by participant survey responses, as illustrated by the following quote regarding stakeholder-based control:

> "I found [stakeholder-based control] to be the easiest method. I was able to focus more on the Mooo of the cow while using this method."

Visual Secondary Task

Control style showed a statistically significant effect on almost all measures ($p < .0001$). No statistically significant effects were observed due to scale or demographic factors. Leader-based and stakeholder-based control performed similarly, and both outperformed predator-based control ($p < 0.005$) with a significant effect size (10%).

Control Activity

The total amount of mouse movement was tracked as a way of measuring the participant's control activity during the scenario. Differences in total mouse movement were significant

Metric	Control Style		
	L vs P	L vs S	P vs S
Total Mouse Movement	$\mathbf{-8,861.83 \pm 1,042.30}$ ($< \mathbf{.0001}$)	$1,878.40 \pm 1,052.67$ (0.1852)	$\mathbf{10,740.00 \pm 1,034.01}$ ($< \mathbf{.0001}$)

Table 5.12: Pairwise Comparisons of control activity over control style. Bold text indicates statistical significance.

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Temporal Demand	0.4138	**0.0011**	0.4656	0.8135	0.5275	0.9372	0.5442
Effort	0.4964	**0.0164**	0.9819	0.5131	0.5682	0.4075	0.3938
Physical Demand	0.9443	**0.0092**	0.0719	0.1316	0.1775	0.8240	0.7123
Frustration	0.0937	$< \mathbf{.0001}$	0.7191	0.7918	0.6520	0.3820	0.8882
Performance	0.1058	**0.0010**	0.2669	0.8341	**0.0288**	0.7014	0.1960
Mental Demand	0.7601	$< \mathbf{.0001}$	0.8736	0.5972	0.4711	0.3382	0.3749
Overall	0.1133	$< \mathbf{.0001}$	0.0714	0.9030	0.4216	0.0339	0.3105

Table 5.13: $Pr > F$ for fixed effects tests on the NASA-TLX survey. Significant differences were found across control style for the overall TLX rating and all subscales. No significant effects were observed across scale. See Table 4.8 for a description of the entries in this table. Bold text indicates statistical significance.

for control style ($p < .0001$), but not scale. The total mouse movement was similar under leader-based and stakeholder-based control, and both were lower than predator-based control by over 50%. Results from survey questions designed to measure control effort (Section 5.3.8) further reinforce these findings.

5.3.7 NASA-TLX

Statistically significant differences were found across control style for the combined NASA-TLX score and all individual components ($p < .01$, see Table 5.13), with the exception of effort, which was weakly significant ($p < 0.02$). No statistically significant differences were found across scale on the combined score or individual components. Robot experience was found to be weakly significant on the performance component, but no other demographic effects were observed. No learning effects were observed, indicating that participants answered consistently across the duration of the experiment.

Metric	Control Style L vs P	L vs S	P vs S
Temporal Demand	$\mathbf{-8.71 \pm 2.32}$ (**0.0013**)	-2.02 ± 2.34 (0.6641)	$\mathbf{6.69 \pm 2.31}$ (**0.0151**)
Effort	-6.40 ± 2.76 (0.0625)	1.30 ± 2.77 (0.8865)	$\mathbf{7.70 \pm 2.76}$ (**0.0200**)
Physical Demand	$\mathbf{-7.47 \pm 3.18}$ (**0.0580**)	2.26 ± 3.18 (0.7581)	$\mathbf{9.73 \pm 3.17}$ (**0.0096**)
Frustration	$\mathbf{-16.21 \pm 2.17}$ (**< .0001**)	-2.11 ± 2.19 (0.6021)	$\mathbf{14.09 \pm 2.16}$ (**< .0001**)
Performance	$\mathbf{-12.16 \pm 3.20}$ (**0.0011**)	-2.75 ± 3.22 (0.6698)	$\mathbf{9.41 \pm 3.19}$ (**0.0132**)
Mental Demand	$\mathbf{-10.85 \pm 2.42}$ (**0.0001**)	1.41 ± 2.43 (0.8325)	$\mathbf{12.25 \pm 2.42}$ (**< .0001**)
Overall	$\mathbf{-12.43 \pm 2.12}$ (**< .0001**)	-0.86 ± 2.05 (0.9068)	$\mathbf{11.56 \pm 2.00}$ (**< .0001**)

Table 5.14: Pairwise comparisons of NASA-TLX survey results across control style. Bold text indicates statistical significance.

Only small differences were found between leader-based and stakeholder-based control, which were not statistically significant. Both control methods showed lower workload than predator-based control ($p < 0.01$) on the temporal demand, mental demand and frustration components, along with the overall score. The effect size was significant, generally 5-15 points on a 100 point scale (Table 5.14). These results match those obtained for other workload measures, including secondary tasks (Section 5.3.6), control activity (Section 5.3.6), and post-scenario surveys (Section 5.3.8). Of special note is the significance of the physical demand component between stakeholder-based and predator-based control, which agrees with our results on neglect time and control activity.

5.3.8 Post-experiment Surveys

All of the post-scenario survey questions had statistically significant differences across control style ($p < .0001$, see Table 5.15). Leader-based and stakeholder-based control scored similarly ($\Delta\mu = 0.1$, $p > .85$), and both scored better than predator-based control ($p < 0.001$). Leader-based and stakeholder-based control reported better control, less difficultly, and less frustration by 1.5 points on a 7 point scale (Table 5.17). These results match our control

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Controllability	0.1550	$<$ **.0001**	0.7687	0.1592	0.1618	0.6439	0.0900
Frustration	**0.0075**	$<$ **.0001**	0.7051	0.1726	0.8571	0.4196	0.2540
Difficulty	**0.0126**	$<$ **.0001**	0.5892	0.2271	0.3215	0.5475	0.0766
Ease of control	0.2153	$<$ **.0001**	0.4800	0.2216	0.4069	0.2523	0.7378

Table 5.15: $Pr > F$ for fixed effects tests on the post-scenario surveys. Bold text indicates statistical significance.

Metric	20 vs 50	20 vs 100	50 vs 100
Frustration	$\mathbf{0.56 \pm 0.17}$	0.27 ± 0.17	-0.29 ± 0.17
	(**0.0052**)	(0.2650)	(0.2070)
Difficulty	$\mathbf{0.51 \pm 0.17}$	0.14 ± 0.17	-0.37 ± 0.17
	(**0.0113**)	(0.6765)	(0.0912)

Table 5.16: Pairwise comparisons of post-scenario surveys across scale. Bold text indicates statistical significance.

Metric	L vs P	L vs S	P vs S
Controllability	1.91 ± 0.20	0.11 ± 0.20	$\mathbf{-1.80 \pm 0.20}$
	($<$ **.0001**)	(0.8511)	($<$ **.0001**)
Frustration	-1.50 ± 0.19	-0.10 ± 0.19	$\mathbf{1.40 \pm 0.19}$
	($<$ **.0001**)	(0.8595)	($<$ **.0001**)
Difficulty	-1.56 ± 0.19	-0.02 ± 0.19	$\mathbf{1.54 \pm 0.19}$
	($<$ **.0001**)	(0.9922)	($<$ **.0001**)
Ease of control	1.90 ± 0.21	-0.03 ± 0.21	$\mathbf{-1.93 \pm 0.21}$
	($<$ **.0001**)	(0.9877)	($<$ **.0001**)

Table 5.17: Pairwise comparisons of post-scenario surveys across control style. Bold text indicates statistical significance.

Metric	Scale	Style	Sex	Vision	Robot	Games	Order
Total Neglect Time	0.6593	$<$ **.0001**	0.5092	0.8823	0.2211	0.4226	0.9015
Longest Neglect Interval	0.0665	$<$ **.0001**	0.8333	0.5404	0.1169	0.6245	0.8570
Average Neglect Time per Interval	0.2529	$<$ **.0001**	0.5612	0.9447	0.2233	0.5996	**0.0419**
Number of Neglect Intervals	**0.0070**	**0.0105**	**0.0147**	0.6164	0.4226	0.2911	**0.0110**

Table 5.18: $Pr > F$ for fixed effects tests on neglect time metrics. Bold text indicates statistical significance.

activity results (Section 5.3.6), secondary task results (Section 5.3.6) and NASA-TLX results (Section 5.3.7). Participants reported slightly higher frustration (0.5 points) with 20 agents than with 50 agents, but no other differences across scale were observed (Table 5.16). No statistically significant demographic differences or learning effects were found, indicating that participants answered consistently over the duration of the experiment.

5.3.9 Neglect Time

In addition to measuring the total amount of mouse movement (Section 5.3.6), we used neglect time measures to examine how the participant's movement varied over time. We found that control style had a statistically significant effect ($p < .001$) on total and average neglect time, along with the maximum neglect interval (Table 5.18). Additionally, the number of neglect intervals was weakly significant at $p < 0.015$. Statistically significant differences were found between all three control styles. Stakeholder-based control showed the most neglect time on all measures, followed by leader-based control, and then predator-based control. The neglect time differences between control strategies were large, with total neglect time differences of 14 and 24 seconds over the 120 second scenario (Table 5.19).

Scale was only found to be significant on a single pairwise comparison of the number of neglect intervals between 20 and 50 agents. More neglect intervals were observed for 50 agents than for 20 agents ($>$ 10% effect size, see Table 5.20), which could indicate a slight difference in control strategies. The number of neglect intervals also showed a weakly significant difference across sex and a weakly significant learning effect, which could also indicate slight differences in control strategy.

	Control Style		
Metric	L vs P	L vs S	P vs S
Total Neglect Time	$\mathbf{24.26 \pm 2.08}$ ($< \mathbf{.0001}$)	$\mathbf{-14.44 \pm 2.09}$ ($< \mathbf{.0001}$)	$\mathbf{-38.70 \pm 2.08}$ ($< \mathbf{.0001}$)
Longest Neglect Interval	$\mathbf{2.21 \pm 0.53}$ ($\mathbf{0.0004}$)	$\mathbf{-3.38 \pm 0.54}$ ($< \mathbf{.0001}$)	$\mathbf{-5.60 \pm 0.53}$ ($< \mathbf{.0001}$)
Average Neglect Time per Interval	$\mathbf{0.23 \pm 0.07}$ ($\mathbf{0.0066}$)	$\mathbf{-0.28 \pm 0.07}$ ($\mathbf{0.0008}$)	$\mathbf{-0.51 \pm 0.07}$ ($< \mathbf{.0001}$)

Table 5.19: Pairwise comparisons of neglect time metrics across control style. Bold text indicates statistical significance.

	Scale		
Metric	20 vs 50	20 vs 100	50 vs 100
Number of Neglect Intervals	-4.06 ± 2.72 (0.3025)	4.97 ± 2.71 (0.1676)	$\mathbf{9.03 \pm 2.73}$ ($\mathbf{0.0049}$)

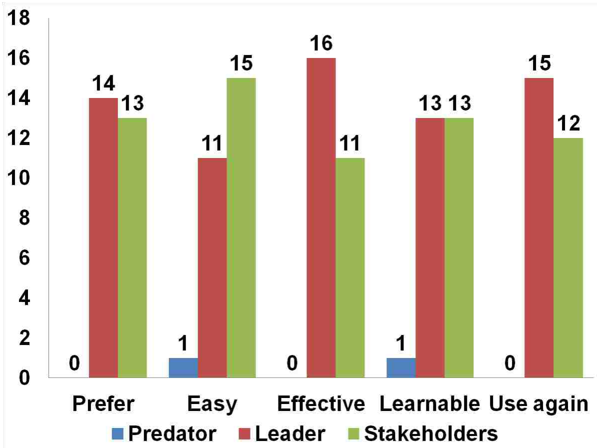
Table 5.20: Pairwise comparisons of neglect time metrics across scale. Bold text indicates statistical significance.

5.4 Discussion

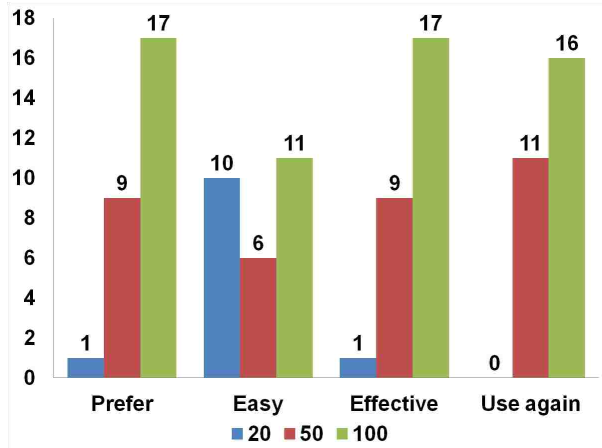
In this section with synthesize the results of the statistical analysis found in Section 5.3 and discuss our findings on the impact of HSI, along with implications for swarm design and future research.

5.4.1 Participant Responses

Participant responses to the post-experiment preference survey (Section 4.3.6) reinforced many of the ideas found in this thesis and provided additional themes and insights, which are supported by our objective results. In this section we summarize the themes that emerged, and discuss how they tie into our research. As each of our results are discussed in turn, we reference these themes and tie in our objective results from the user study. The themes discussed in this section provide a framework for understanding our results and provide direction for future research (see Appendix A).



(a) Control style



(b) Scale

Figure 5.1: Participant preference survey

Participant Preference Survey

At the conclusion of the experiment, participants were asked several questions about their preferred control style and number of agents (see Table C.1 for the list of questions). Participants preferred leader or stakeholder-based control to predator-based control on all of the survey questions (Figure 5.1). Leader-based control was rated more effective, while stakeholder-based control was rated as easier to learn, matching our results for swarm performance and neglect tolerance. Participants preferred larger swarms on most of the questions and considered group sizes of 20 and 100 equally easy to learn. We hypothesize that the preference for larger groups is to due to better performance (Section 5.4.2) and better heading stability of the group (Section 5.4.7), along with dropout tolerance and adjustable autonomy, which we discuss in this section. The survey results match our objective measures and show that participant's perception of the swarm's performance matched our objective measures.

Adjustable Autonomy

Due to the diminishing performance returns of larger swarm sizes, many participants stated that they adapted their strategy based on the size of the swarm. They stated that they actively tried to keep small groups together, but allowed some fragmentation in larger groups[2]. In fact, some of the participants said that they deliberately chose to let some of the agents wander to pick up additional points, as illustrated by the quote below. Thus, as the size of the group increased beyond what was needed to accomplish the task, participants allowed more agents to function autonomously, while directing the rest of the group. Some participants also further split the group into subgroups. These results may indicate that suitably designed swarms have some inherent degree of adjustable autonomy, allowing the human operators adjust their interaction effort, or alter the autonomy level of a portion of the swarm based on the current situation and their workload. Interestingly, participants in the first small-scale user study also deliberately fragmented the group, which matches our large-scale results (see Section 3.4.4).

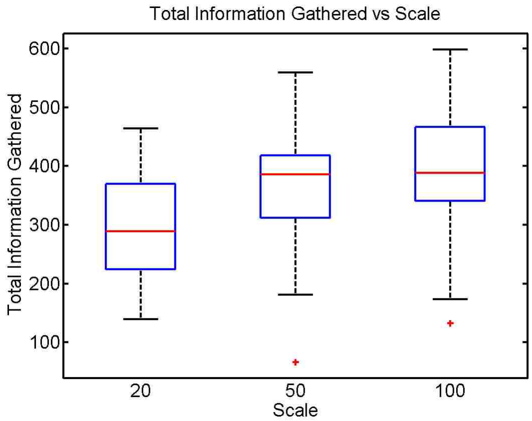
The following quote illustrates how participants made use of this idea during the study:

> "[My strategy for large groups was to] move slowly to keep control of [the robots], or allow some to diverge from the group, purposely allowing them to move freely, thus creating an opportunity for them to randomly hit barrels."

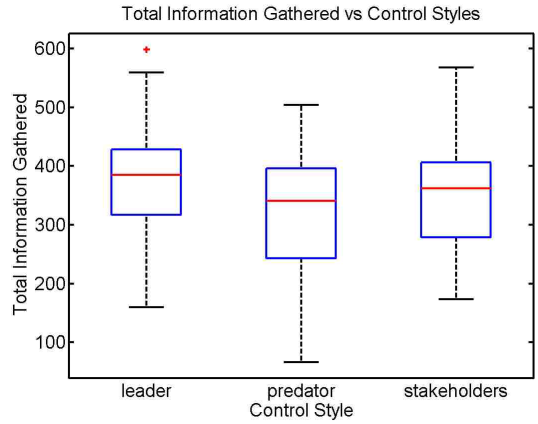
Swarm Affordances

Participants' comments reinforced our previous hypothesis on the affordances offered by the control methods we evaluated. Most of the responses for predator-based control described fragmentation and some strategies related to it, compared to very few references to fragmentation for the other control methods. Participants either adapted strategies to avoid it, such as patrolling the perimeter of the swarm, or attempted to make use of it by deliberatively fragmenting the group. The comments here, along with our prior findings and objective results, reinforce the idea that predator-based control affords fragmentation.

[2]Eighteen of twenty-seven participants said they adapted their strategy based on group size. Nine participants said they deliberately fragmented the group and additional nine participants said they were less concerned about fragmentation with larger group sizes.



(a) Information gathered vs scale



(b) Information gathered vs control style

Figure 5.2: The total information gathered as a function of (a) scale and (b) control style. The performance of the swarm was affected by both number of robots and control style ($p < .001$). Performance increased with swarm size up to a saturation point. Control style also significantly affected swarm performance.

Results were similar for leader and stakeholder-based control. The responses for leader-based control emphasized going slow, staying close to the group, and keeping some swarm members within the radius of influence, which all indicate sustained influence. Similarly, responses for stakeholder-based control described it as easier, and referenced neglect time with phrases such as "just put the target on the barrel" or "I was able to put the cursor on the target and focus on [the secondary tasks]." The survey and participant comments for stakeholder-based control are consistent with our findings in Section 5.3.9.

5.4.2 Performance

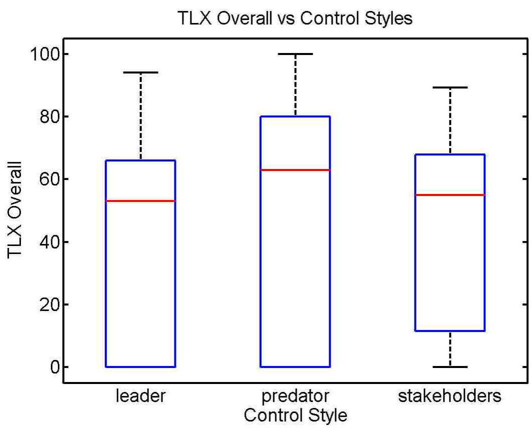
We found that both scale and control style significantly impacted swarm performance (Figure 5.2). Larger swarms scored higher on all of our performance measures, but showed diminishing returns as more agents were added, representative of many real-world tasks. Even with a simple swarm algorithm, swarm performance increased simply by adding more agents, without altering inter-agent dynamics, changing the human-operator's control problem, requiring additional knowledge shared among agents, allocating additional bandwidth, or most notably, increasing operator workload. Additionally, we found evidence that the distributed nature of the swarm allows for an implicit degree of adjustable autonomy that a human operator can manage on-the-fly. This is accomplished by splitting off individuals or dividing the swarm into groups as determined by the task, operator workload, and swarm performance.

We also found that choices made in the design of the swarm, such as control style, have a significant impact on the performance of the swarm, which illustrates the need to perform further research into the building blocks of swarm systems and the effect that they have on the human-swarm system, including the individual agents, group behaviors, and the workload of the human operator. We found that different control methods not only affected performance, but also provided different affordances to the operator. For example, we found that predator-based control was outperformed by other control methods, but afforded richer behaviors that could increase performance on some tasks.

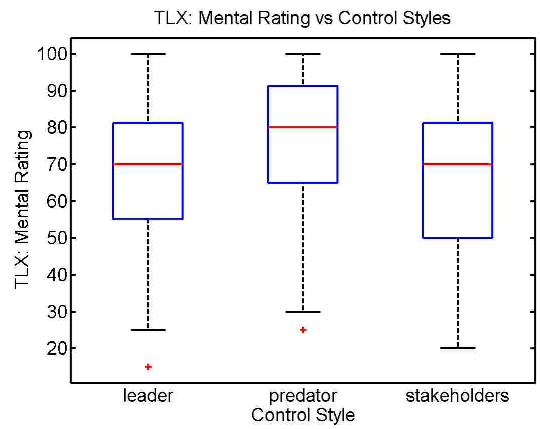
5.4.3 Workload

The primary goal of the large-scale user study was to evaluate the effect of control style and scale on the workload of the human operator while controlling a swarm. We evaluated the participant's mental workload using a variety of objective and subjective methods, including multiple metrics on two secondary tasks, NASA-TLX, post-scenario surveys, and post-experiment preference surveys. There was strong agreement between methods and the results showed that operator workload was affected by control style, but not scale. The control style results were strongly signficiant ($p < .001$) across all measures using both fixed effect and pairwise tests. Scale was only significant on a single pairwise test[3], and most measures were far from significant ($p > .10$ or more). Furthermore, the effect size across scale on all measures was small and at most of limited practical significance had there been significance. We also evaluated a variety of demographic and confounding factors and did not find any significant results. Based on these results, we have confidence that we gathered sufficient data and sufficiently accounted for any confounding factors.

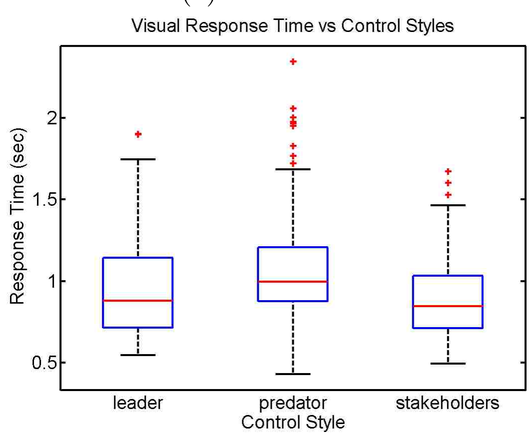
[3]There was a difference of one half survey tick between 20 and 50 agents on the frustration component of NASA-TLX



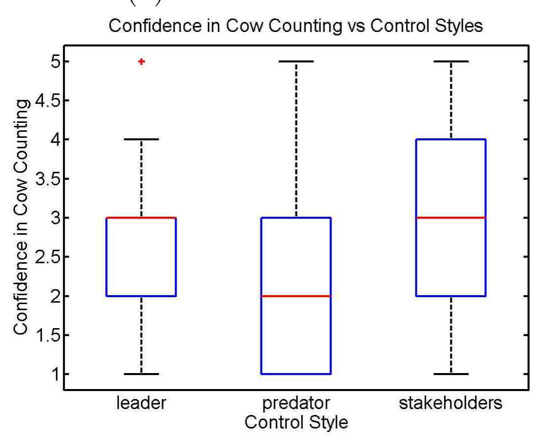
(a) TLX overall



(b) TLX mental demand



(c) Response time on visual task



(d) Auditory Task Confidence

Figure 5.3: Mental workload across control styles

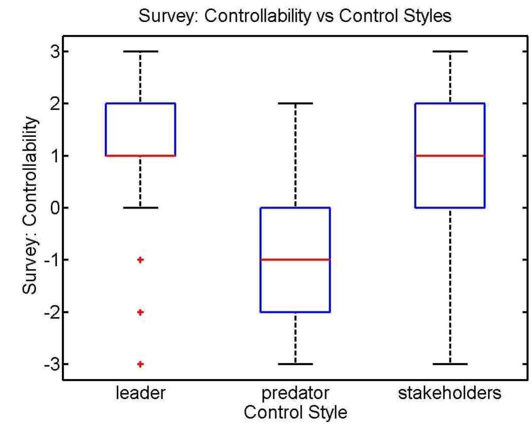
We found that swarm performance increased with scale, but mental workload did not. In contrast to a traditional supervisory control approach, swarm management allowed the human operator to manage additional robots and increase performance without the need to divide attention between robots and incur additional workload. Additionally, the results of the survey indicate that properly design swarms may afford inherent adjustable autonomy, allowing the human operator to scale his or her control activity based on his or her workload at the time. These results show that swarm management is a promising approach for managing large numbers of robots.

The results of the user study also show that swarm design choices, in this case control style, can impact the workload of the human operator while managing the swarm. We establish relative workload between three simple control styles, leader, predator, and stakeholders. Leader-based and stakeholder-based control have the lowest mental workload and require the least control activity. Our results from participant surveys reinforce our earlier findings on control method afforances. Predator-based control afforded the richest behaviors in our user studies, followed by leader, and then stakeholders, which also matches our neglect time results. This suggests a trade-off between mental workload and neglect tolerance, and expressiveness[4]. Furthermore, the local influence of the control methods and decentralized nature of the swarm potentially allows the control method to be altered on-the-fly to adapt to the current situation facing the operator. These ideas and results illuminate many exciting possibilities for further research on swarm control.

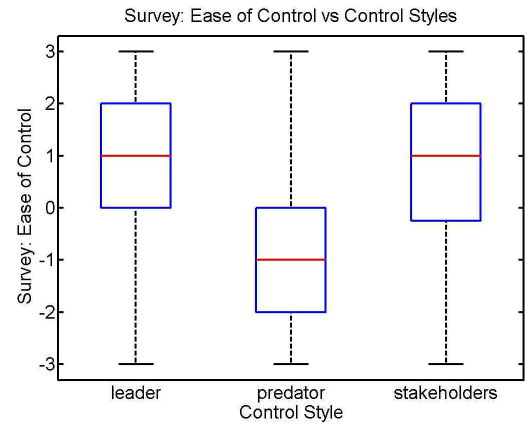
5.4.4 Control Activity

In control theory, *control effort* is the amount of energy put into a system to move it toward a desired state. In the context of HRI, we define the operator's control activity as the amount of work an operator must perform to reach a desired state of the human-robot system. In addition to mental workload, control activity provides another component of the overall

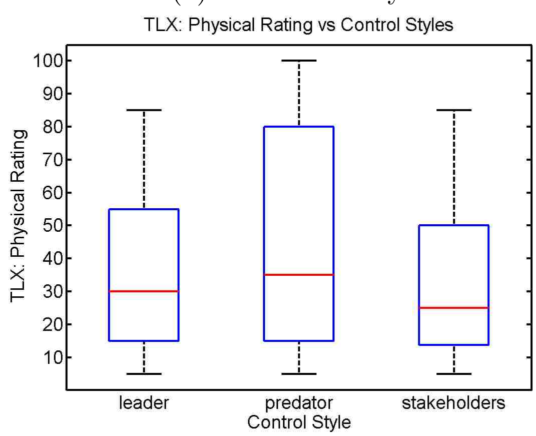
[4]See Appendix A for a more extensive discussion and hypothesis for future research.



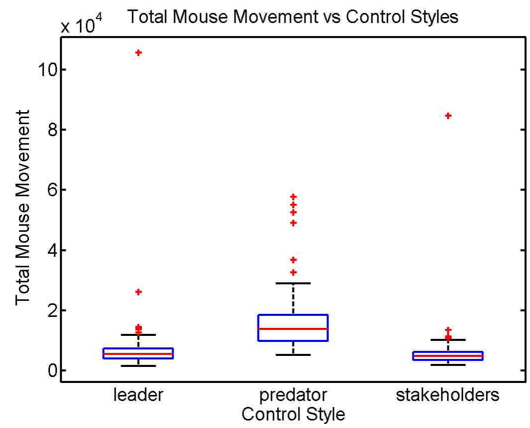
(a) Controllability



(b) Ease of control survey rating



(c) TLX physical effort

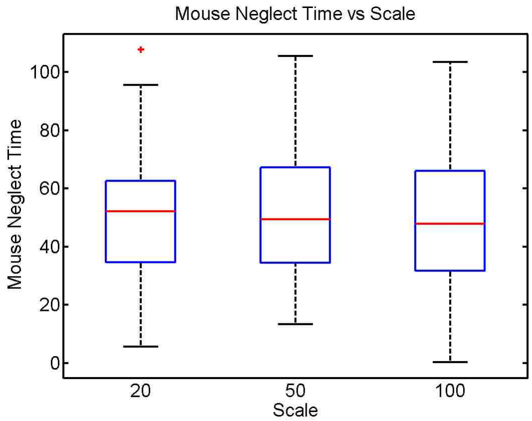


(d) Total mouse movement

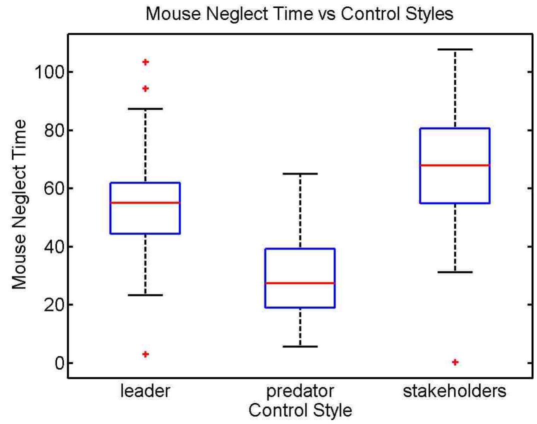
Figure 5.4: Control activity

workload of the human operator. To measure control activity, we primarily considered the total movement of the participant's mouse during the scenario, which showed statistically significant differences across control style. Additionally, several related measures, including TLX physical effort, and both survey questions related to control, showed statistically significant effects that matched our mouse movement results.

We found that predator-based control required the most control activity across all measures (Figure 5.4). Leader-based and stakeholders-based control required similar amounts of control activity, both much lower than predator-based control. When combined with the



(a) Scale



(b) Control style

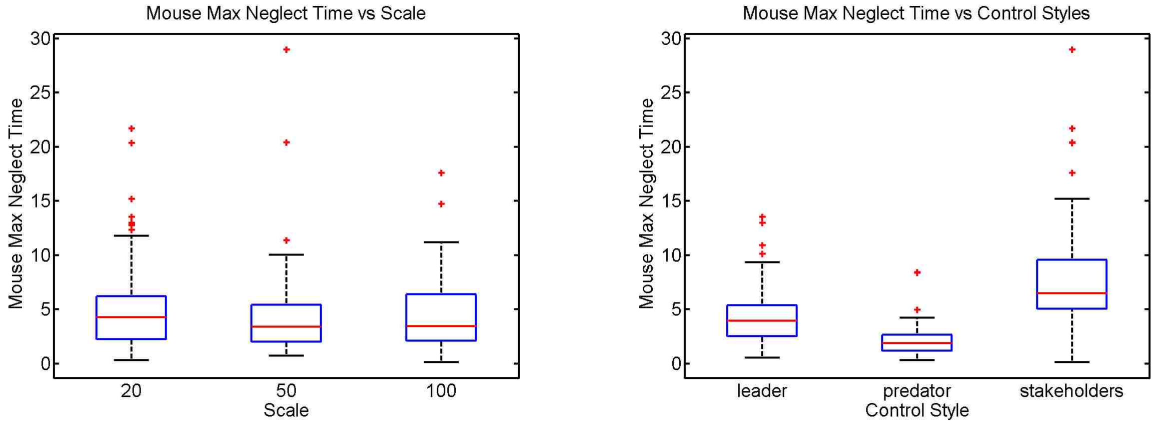
Figure 5.5: Total neglect time

mental workload results shown in Section 5.4.3, these results show an overall difference in the operator's workload across control style.

5.4.5 Neglect Time

We examined not only the total amount of effort expended by the human operator but also how that effort was expended with regard to time using the concept of neglect time. Neglect tolerance of the swarm is a desirable characteristic, which allows the human operator time to focus on other tasks or potentially exert control over multiple swarms. In this experiment, we measured neglect time as periods when the participant's mouse remained stationary; e.g. the control signal to operator-controlled agents in the swarm was constant and the control activity was zero.

Our analysis showed substantial differences in the total neglect time, average neglect time per interval, and the longest neglect interval (see Figure 5.5 and 5.6). Stakeholder-based control showed the most neglect time, followed by leader-based control, and then predator-based control. Notably, neglect time is the only group of metrics that showed statistically significant differences between leader-based and stakeholder-based control. These differences were also referenced in participant comments, as discussed in Section 5.4.1. Based on these



(a) Scale

(b) Control style

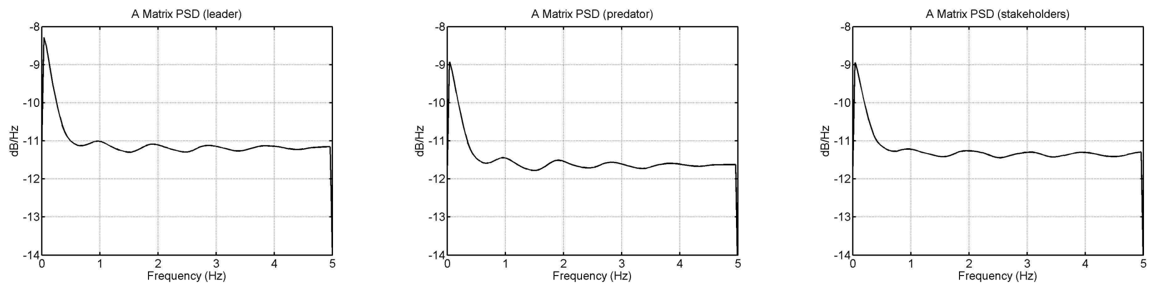
Figure 5.6: Longest neglect interval

results, we hypothesize that neglect tolerance is another attribute afforded by certain design choices, and that stakeholder-based control affords neglect tolerance better than other control methods evaluated in this user study.

5.4.6 Effects on Group Topology

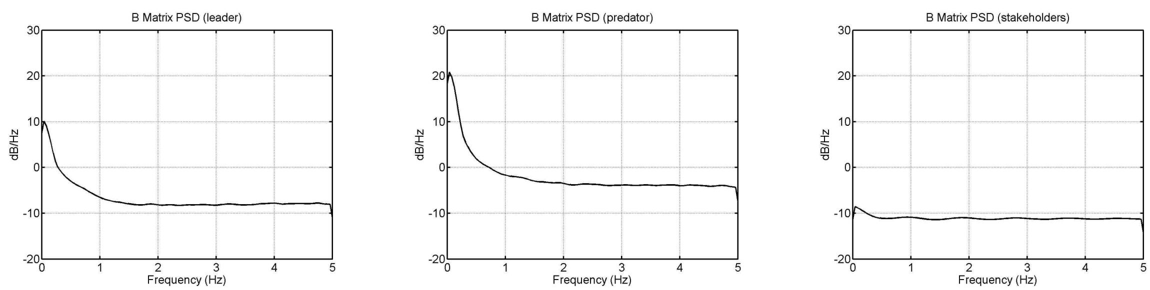
Using the methods described in Chapter 3, we analyzed the effect of HSI on the topology of the swarm. The results from the large-scale user study matched our previous results (Section 3.4) and reinforced our previous findings on the affordances provided by different control methods (see Section 3.3.4).

We found that leader-based control sustained influence better than predator-based control (Figure 5.8), but caused more scrambling within the swarm (Figure 5.7), reproducing our previous results found in Section 3.4.5). The large-scale study also addressed the previous limitation due to the use of a switching controller ($w_p = \infty$) for leader and predator influence. We found that the non-switching controller did not alter the ordering of the results, but did substantially reduce the effect size for scrambling within the group. Using a non-switching controller, the differences in disruptions in group topology between control methods are



(a) Leader (b) Predator (c) Stakeholders

Figure 5.7: A Matrix PSD for the large-scale user study using a non-switching controller ($w_p < \infty$). Predator based-control caused the least disruption to the group, followed by leader-based and stakeholder-based control. However, the non-switching controller reduced to the effect size to $< 1dB$, so the differences are not of practical significance. Compare to Figure 3.10d.

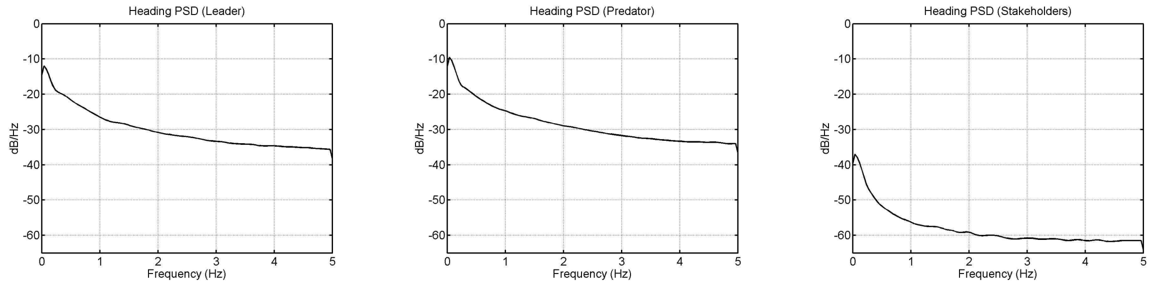


(a) Leader (b) Predator (c) Stakeholders

Figure 5.8: B Matrix PSD for the large-scale user study using a non-switching controller ($w_p < \infty$). Stakeholder-based control sustained influence best by a wide margin, followed by leader-based control, and then predator based-control. Compare to Figure 3.7d

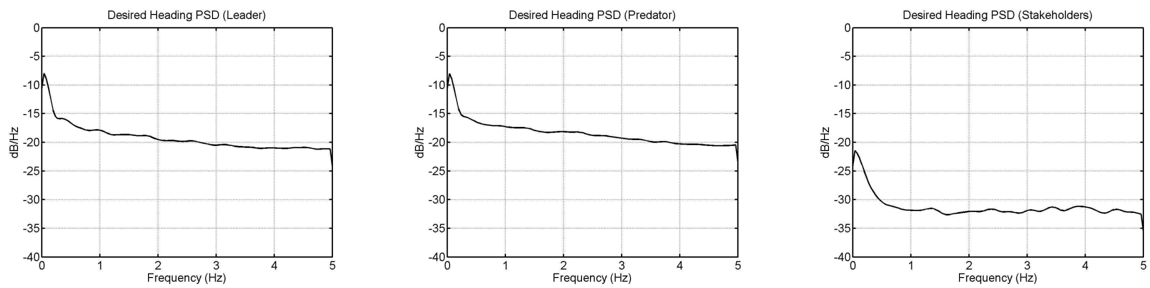
$< 1dB$. This means that there is little practical difference in disruption to group topology across control methods, provided that a non-switching controller is used.

The large-scale user study also included stakeholder-based control, which was not evaluated in our previous results. We found that stakeholder-based control sustained influence better than leader-based or predator-based control by a wide margin (20dB and 30dB respectively). Stakeholder-based control also caused less disruption to group topology than leader-based control, but more than predator-based control, though the difference is not of practical significance. Based on participant feedback (Section 5.4.1) and our neglect time results (Section 5.4.5), we hypothesize that stakeholder-based control affords both sustained



(a) Leader (b) Predator (c) Stakeholders

Figure 5.9: PSD of the changes in heading of participant-controlled agents



(a) Leader (b) Predator (c) Stakeholders

Figure 5.10: PSD of the changes in desired heading of participant-controlled agents

influence and neglect tolerance, just as leader and predator control afford sustained influence and fragmentation respectively.

5.4.7 Effects on Individual Agents

In addition to examining the effect of HSI on the topology of the swarm, we also examined the effect on the individual agents. To conduct this analysis, we computed power spectral density plots of changes position, heading, and desired heading over time, for both swarm members and participant-controlled agents. Leader-based and predator-based control showed similar variations in desired heading, while predator-based control showed more variation in desired heading and position. Stakeholder-based control showed the least variation in position, heading, and desired heading by a substantial margin. We also found a decrease across scale in the variation of heading and desired heading, but attribute these differences, at least partially, to changes in participant strategy across scale, as described in Section 5.4.1.

5.5 Conclusion

The results from the user study confirm our experiment hypothesis and demonstrate the utility of swarm systems for controlling large numbers of robots. In traditional supervisory control, the workload of the human operator increases with the number of robots managed. Our results show that properly designed swarm systems can increase performance without increasing workload. This result has significant implications for the future design of swarm systems and the use of decentralized swarms to manage massive numbers of robots. Additionally, we also found evidence indicating that well-designed swarm systems possess an inherent degree of adjustable autonomy, allowing additional flexibility for the human operator.

The results in this chapter also reinforce our previous results on the affordances of different control styles and their impact on the topology of the swarm over time. We further find that control style significantly impacts performance, operator workload, control activity, and neglect time. Together, these results form the idea of control complexity, which represents a tradeoff between the expressiveness of a given control style and the workload of the human operator(s). Appendix A discusses this idea in detail as a hypothesis for future work.

The analysis of the data also illustrates the need to identify and quantify the basic building blocks of swarm algorithms and determine their impact the human-swarm system. Our analysis shows that there not only exist advantages and disadvantages but complex tradeoffs between components that must be considered in the context of the swarm operator and mission environment.

Chapter 6

Conclusion and Future Work

The emerging field of Human-Swarm Interaction (HSI) deals with the effective management of swarms by human operators. In this thesis we have contributed foundational work on the effect of HSI (a) on the individual robots, (b) on the group as a whole, and (c) on the workload of the human operator. We have (1) shown that existing general swarm algorithms are feasible on existing robots and can display collective behaviors as shown in simulations in the literature, (2) analyzed the effect of interaction style and neighborhood type on the swarm's topology, (3) demonstrated that operator workload stays stable as the size of the swarm increases, and (4) found that operator workload is influenced by the interaction style. As part of this work, we have also contributed results on the affordances and tradeoffs of three general control methods that can be applied to a variety of swarm algorithms.

In this chapter, we summarize the contributions and results of this thesis and present opportunities for continued research based on our work. We also direct the interested reader to Appendix A, which presents the unifying theme of control complexity as a hypothesis for future work.

6.1 Summary

Robots, including UAVs, have found increasing use in helping humans with dangerous and difficult tasks [61]. The number of robots in use is increasing and is likely to continue increasing in the future. As the number of robots increases, human operators will need to coordinate and control the actions of large teams of robots. While multi-robot supervisory control has been widely studied, it requires that an operator divide his or her attention between robots. Consequently, the use of multi-robot supervisory control is limited by the number of robots that a human or team of humans can reasonably control [57]. Swarm robotics – large numbers of low-cost robots displaying collective behaviors – offers an alternative approach by providing the operator with a small set of inputs and parameters that alter the behavior of a large number of autonomous or semi-autonomous robots. Researchers have asserted that this approach is more scalable and offers greater promise for managing huge numbers of robots [61].

The objective of our research was to evaluate the scalability of swarm systems in terms of operator workload and evaluate the effect of HSI on the human-swarm system. Specifically, we have evaluated the effect of control style and scale (a) on the individual members of the swarm, (b) on the topology of the swarm itself, and (c) on the workload of the human operator. Along the way our research has also yielded important results on control method affordances, swarm neglect tolerance, the effects of neighborhood definition on swarm topology, and adjustable autonomy within swarms. We now summarize each of our contributions, chapter by chapter, and present the overall results and impact of this work.

We begin in Chapter 1 by defining criteria for the selection of a swarm model and control methods to ensure our results are applicable and can be generalized to a broad set of robot hardware and situations. We select Couzin's model and three general control methods (leader, predator, and stakeholders), which we integrate into the model. We subsequently incorporate additional extensions to Couzin's model for bounding, obstacle avoidance, and error modeling as required by our research.

In Chapter 2, we demonstrate that the models and scenarios used in our user studies are feasible and can be implemented on real hardware. We select robot hardware and demonstrate both proximate and remote interaction in a simulated force protection scenario; using only the hardware and sensing capabilities of the robots. Next, we demonstrate how an abstract swarm algorithm can be successfully implemented on commercially available robots. We also provide relevant considerations encountered during our experiments. Further experiments show that our augmented version of Couzin's model is sufficient to navigate an indoor environment using onboard sensors. We then evaluate the physical parameters and localization error of the robots and use them to construct an accurate model. Finally, we validate this model and use it to show that physical robots are capable of displaying the collective behaviors demonstrated by Couzin in simulation.

In Chapter 3, we summarize our previously published results which have identified important affordances and tradeoffs based on interaction style and neighborhood definition. We find that: First, using a topological rather than metric neighborhood decreases the amount of scrambling within the group. Second, parameter-based control performs well and requires little management, but must be optimized for the scenario *a priori*. Third, leader-based control sustains influence over the group, but the sustained influence causes more scrambling within the group than predator-based control. Fourth, predator-based control provides expressive behaviors such as splitting of the group and interacts with fewer members of the swarm for a shorter duration per agent. However, predator-based control does not sustain influence and the expressive behaviors it affords require additional attention on the part of the human operator.

In Chapters 4 and 5 we present the design, implementation, and analysis of the culminating user study of this thesis. We first define the goals and objectives of the user study, and then define the metrics used to evaluate these objectives, along with changes to the model as required by the study. Following the study, we conduct an extensive statistical analysis of the data and find that the results from the user study confirm our experiment hypothesis and demonstrate the utility of swarm systems for controlling large numbers of robots. Our results show that properly designed swarm systems can increase performance without increasing workload as more agents are added. The results also reinforce our previous findings on the affordances of different control styles and their impact on the topology of the swarm over time. We further find that control style represents a tradeoff that significantly impacts performance, operator workload, control activity, and neglect time in addition to the affordances it provides. Additionally, we find evidence indicating that well-designed swarm systems possess an inherent degree of adjustable autonomy, allowing additional flexibility for the human operator.

Overall, two major themes emerged from the results of our research: (1) a validation of the feasibility and scalability of human-swarm systems incorporating robot hardware, algorithms, HSI, and operator workload, and (2) an evaluation and analysis of the affordances and tradeoffs provided by swarm design choices, which primarily focused on control method.

With regard to theme (1), swarms provide several desirable characteristics, including robustness and the ability to scale to a large number of robots. In our research, we found that swarms not only scaled algorithmically, but also in terms of performance and operator workload. Additionally, we found that the decentralized nature of properly designed swarm systems may support an implicit degree of adjustable autonomy; which permits human operators to dynamically scale their control effort based on current workload and task demands. These results were obtained in a full user study, backed by rigorous analysis, simulations, modeling, and hardware validation. Based on our validation, experiments, and analysis, we found large-scale HSI to be both feasible and scalable and conclude that HSI holds promise for managing large numbers of robots.

With regard to theme (2), our experiments also demonstrated that design choices, such as control method or neighborhood definition, have a substantial impact not only on swarm performance, but on all aspects of the human-swarm system. Our evaluation of control methods found relatively large statistically significant differences ($> 10\%$) across operator workload, neglect time, control activity, sustained influence, and the number and duration of interactions. Furthermore, each control method showed a unique set of affordances, which required different control strategies on the part of the human operator. We also found that neighborhood definition made a significant difference in the amount of scrambling within the group. Design choices are not limited to algorithmic considerations, but also include robot hardware and sensors, along with consideration of the task, environment, and human operators involved.

These results illustrate the importance of understanding the impact and tradeoffs between design primitives in order to engineer reliable human-swarm systems. Ultimately, the best choice of control method may depend on considerations for the scenario or task. The small control effort required by parameter-based control may be desirable for controlling massive numbers of agents, while the higher control effort required by the predator-based control may help the operator stay focused and vigilant during long-duration missions.

In conclusion, our results provide a foundational contribution to the emerging field of human-swarm interaction. Our work has demonstrated the scalability and feasibility of swarm systems from an HSI perspective. We have also provided novel results on swarm neglect tolerance, the effects of control style on the swarm and operator workload, relative performance across control style, sustained influence, and the effects of neighborhood definition on group topology. Our work has illustrated not only the relative differences between design choices, but also the importance of understanding the tradeoffs involved. We hope that future work will incorporate and build on our results as it progresses toward a better understanding of swarm systems, HSI, and swarm engineering.

6.2 Directions for Future Work

The research presented in this thesis has provided a foundational contribution to the understanding of the affordances of different design choices and their effect on the human-swarm system. As part of this work, we have identified several related areas that could benefit from additional research, which we describe here. We also discuss areas of this thesis that future work could expand or strengthen. We organize this section into related categories that roughly correspond to the chapters of this thesis.

6.2.1 Swarm Algorithms and Control Methods

In this thesis we selected and evaluated a single swarm algorithm and several general control methods. While care was taken to provide results that generalize as much as possible, the number of algorithms and control methods in the literature is extensive and warrants further investigation. Future work should generalize our results to more algorithms and control methods, with the goal of building a library of well-understood behavioral primitives that can be used to construct desired swarm behavior. Of special interest is the work of McLurkin [49], which defines and implements a variety of simple behaviors of swarm robots.

6.2.2 Physical Robots and Sensor Considerations

As discussed in Section 2.2, the physical robot platform used for the swarm has an important impact on the human-swarm system that merits consideration. In our research, physical robots were primarily used for validation and calibration of model parameters for the user study. Future work should investigate the effect of a broader range of parameters, hardware configurations, and sensing capabilities on the human-swarm system. Of special interest is the tolerance of existing swarm algorithms to localization error, sensor noise, bandwidth limitations, and communication dropouts. This is especially interesting in the case of *sensor-based resolution*, where each robot's sensors directly limit the quantity and quality of information obtained about other robots. In the context of Couzin's model, this may include investigating the minimum value of α before swarm fragmentation occurs.

The scope of our research was also limited by the number of physical robots available. Consequently, we conducted our validation of group behaviors and user studies in simulation, using the parameters and capabilities of the physical robots. Future work should replicate our results on physical robots. Because people react differently to robots that are physically present [6], a user study incorporating proximate HSI with physical robots would provide a valuable compliment to our remote HSI user studies.

6.2.3 HSI

Human-swarm interaction is an emerging research field with a variety of open problems and opportunities for foundational research. Our results not only contribute to the field, but also provide a basis on which other research may be built. Future work should continue to evaluate effects of swarm design choices on the human-swarm system. This includes additional control methods, neighborhood definitions, swarm algorithms, and parameters, among others. Future work should also consider more dynamic control methods, such as switching between control primitives, collective structures, and transitions between collective structures. In addition to more complex control, increased agent complexity should also be considered. This may take the form of increased or adjustable autonomy, more complex algorithms, colony-like behaviors, or heterogeneous swarms. Future work should also consider design choices that directly impact the human operator, such as partial observability or multi-operator control.

6.2.4 Control Complexity

In our analysis of participant comments in Section 5.4.1, we identified a unifying theme incorporating our results on swarm affordances, neglect time, and operator workload. Participants consistently described behaviors and control strategies of different complexity for each of the control methods in the study. We found that control methods associated with more complex strategies provided richer affordances, but at a cost of higher workload and decreased neglect time. We term this tradeoff *control complexity*, which we describe in detail in Appendix A. Control complexity provides a possible unifying theme for HSI control methods and may serve as a useful hypothesis in future work.

Appendix A

Swarm Control Complexity: a Hypothesis for Future Work

A.1 Introduction

Participant comments provided a single unifying theme that ties together our results on mental workload, neglect time, and control method affordances. We term this idea *control complexity* and define it as the behavioral richness and degree of planning afforded by different swarm control methods. Control complexity also represents a tradeoff between the expressiveness of a given control method and the operator's mental workload, as shown by the results of the study. Increasingly complex control methods afford richer behaviors, but come at the cost of increased mental workload and decreased neglect time. This is especially interesting for swarm robot systems, where the control method can potentially be changed on-the-fly. We further hypothesize that the three control methods used in the user study represent three different levels of control complexity, which we term *direct influence*, *indirect influence* and *strategic influence*, as illustrated by the following participant comment regarding stakeholder-based control:

> "[Stakeholders] was the easiest to use because I didn't have to think about moving something to move another object. I felt like I was moving the team itself."

A.2 Levels of Control Complexity

We identified three levels of control complexity in our research: direct influence, indirect influence, and strategic influence. We hypothesize that other levels of control complexity may also exist, such as as full autonomy or pack-based strategies that involve the coordination of multiple operator-controlled agents.

A.2.1 Direct Influence

Direct influence consists of commands issued by the human operator which directly influence the behavior of members of the swarm. The swarm members are not under the direct control of the human operator, but receive commands, such as parameter changes or goal locations, which alter their behavior. In our experiments, this complexity level was represented by stakeholder-based control. The parameter-based control described in Section 3.2 also falls into under this category, as does work by Haas [39], Fields et al. [28] and Conradt et al. [18]. Participants in the user study described stakeholder-based control using passive verbs and discrete events, such as "put [or hold] the target on the barrel" or "get the group's attention, then move to the barrels." Our results also showed the direct influence showed low mental workload and high neglect time, but afforded only simple behaviors.

A.2.2 Indirect Influence

Indirect Influence represents a moderate level of control complexity in which the human operator(s) directly control one or more agents, which swarm members recognize as distinct, that are used to influence the swarm. In our study, indirect influence was represented by leader-based control. Participants directly controlled the leader, which influenced members of the swarm when they were in range. Participants described this complexity level using continuous actions and more active verbs, such as "lead", "stay close", or "keep agents inside the circle." Our results showed sustained influence, low workload, and moderate neglect time. Other examples of indirect influence include work by Bashyal et al. [9] and firefighter interaction in the GUARDIANS project [59].

A.2.3 Strategic Influence

The highest level of control complexity represented in our study is *strategic influence* in which the operator executes a strategy on one or more distinct agents, which in turn influence the swarm. In our study, this was represented by predator-based control, which participants described in terms of strategies executed on the predator, such as "circling," "patrolling," or "splitting the group." Additionally, participants described multiple distinct strategies, whereas comments for lower levels of complexity were very similar to one other. Strategic influence afforded the richest set of behaviors, but came at the cost of higher workload and lower neglect time.

A.3 Conclusion

Control complexity represents a novel taxonomy for HSI methods found in the literature. It presents a unifying theme that ties together our objective results for operator workload, swarm performance, and swarm neglect tolerance, as well as incorporating our participant comments and our work on swarm affordances. In this thesis, we identify three levels of control complexity: direct influence, indirect influence, and strategic influence represented by stakeholders, leaders, and predators respectively. We show how these levels differ in terms of affordances, swarm performance, and impact on the human operator. We also provide examples of how HSI methods found in prior work fit into these levels. The idea of control complexity represents a novel contribution to the field of HSI which, to our knowledge, has not been explored in the HSI or HRI literature. It is our hope that this idea will provide a hypothesis and unifying theme for future work in the field of HSI.

Appendix B

IRB Informed Consent and Advertisement Flyer

Figure B.1: Flyer advertising the user study

Consent to be a Research Subject

The goal of the Technology-Supported Human-Robot Interaction project is to improve understanding of human-robot interaction. Research subjects will be asked to perform simple information processing while controlling a robot or robot arm. The purpose is to characterize different robot control modes. It is not important that you have extensive computing experience, but we do expect that you have used a joystick and a mouse before. You will also be expected to be able to answer simple arithmetic problems.

You will be asked to perform an arithmetic task, a spatial reasoning task, or a memory-recall task. This test is not used to assess your capabilities to do arithmetic but rather to give you experience in the type of task you will be performing; the test also allows us to calibrate the task to your abilities. In addition to the arithmetic test, you will be asked to control one or more robots/arms via a joystick, mouse, steering wheel, PDA, or microphone. After letting you practice controlling the robot(s)/arm(s) with the joystick, mouse, steering wheel, PDA or microphone in this mode, you will be asked to control the robot(s) while solving arithmetic, spatial reasoning, or memory-recall problems. This process will then be repeated for a second, third, or fourth robot/arm control mode.

Your performance while controlling the robot may be videotaped. Your name and any identifying information about you will be kept in confidence. Selected portions of the video tape may, however, be used when presenting the research results. If such video is used, we will remove all information that would identify you personally and we will attempt to use only video shots that do not include your face.

You may also be requested to describe your experience in using the system. If you are asked to describe your experience, your comments will be transcribed and may be analyzed to assess the usability of the system.

Participation in this study is voluntary. There are no threatening or painful aspects to these experiments. There may be some discomfort for those who are sensitive to motion sickness. If this happens, you can stop the experiment. You may terminate your involvement at any time. The experiment will last no more than 1 hour. You may be requested to return at a later date for an additional experiment of up to 1 hour, but such participation is at your discretion. You will receive $12.00 for your participation for each experiment.

If you have any questions concerning this study you may contact one of the project directors Dr. Michael A. Goodrich (801) 422-6468, mike@cs.byu.edu; Dr. Bryan Morse, (801) 422-8146, morse@cs.byu.edu.p or Dr. Mark Clton, (801) 422-6303, colton@byu.edu. If you have questions regarding your rights as a participant in a research project you may contact BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT 84602, (801) 422-1461, irb@byu.edu.

I ___________________________________ have read the above and consent to
(Please print your name)

participate in this research study.

____ ____________________
Signature Date

Brigham Young University IRB
APPROVED EXPIRES
MAR 14 2012 APR 12 2013

Figure B.2: IRB informed consent form ($12 payment)

Consent to be a Research Subject

The goal of the Technology-Supported Human-Robot Interaction project is to improve understanding of human-robot interaction. Research subjects will be asked to perform simple information processing while controlling a robot or robot arm. The purpose is to characterize different robot control modes. It is not important that you have extensive computing experience, but we do expect that you have used a joystick and a mouse before. You will also be expected to be able to answer simple arithmetic problems.

You will be asked to perform an arithmetic task, a spatial reasoning task, or a memory-recall task. This test is not used to assess your capabilities to do arithmetic but rather to give you experience in the type of task you will be performing; the test also allows us to calibrate the task to your abilities. In addition to the arithmetic test, you will be asked to control one or more robots/arms via a joystick, mouse, steering wheel, PDA, or microphone. After letting you practice controlling the robot(s)/arm(s) with the joystick, mouse, steering wheel, PDA or microphone in this mode, you will be asked to control the robot(s) while solving arithmetic, spatial reasoning, or memory-recall problems. This process will then be repeated for a second, third, or fourth robot/arm control mode.

Your performance while controlling the robot may be videotaped. Your name and any identifying information about you will be kept in confidence. Selected portions of the video tape may, however, be used when presenting the research results. If such video is used, we will remove all information that would identify you personally and we will attempt to use only video shots that do not include your face.

You may also be requested to describe your experience in using the system. If you are asked to describe your experience, your comments will be transcribed and may be analyzed to assess the usability of the system.

Participation in this study is voluntary. There are no threatening or painful aspects to these experiments. There may be some discomfort for those who are sensitive to motion sickness. If this happens, you can stop the experiment. You may terminate your involvement at any time. The experiment will last no more than 1 hour. You may be requested to return at a later date for an additional experiment of up to 1 hour, but such participation is at your discretion. You will not be paid for participating in this experiment; participation is strictly voluntary.

If you have any questions concerning this study you may contact one of the project directors Dr. Michael A. Goodrich (801) 422-6468, mike@cs.byu.edu; Dr. Bryan Morse, (801) 422-8146, morse@cs.byu.edu.p or Dr. Mark Clton, (801) 422-6303, colton@byu.edu. If you have questions regarding your rights as a participant in a research project you may contact BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT 84602, (801) 422-1461, irb@byu.edu.

I ___________________________________ have read the above and consent to
(Please print your name)

participate in this research study.

____ ________________

Signature Date

Brigham Young University IRB
APPROVED EXPIRES
MAR 1 4 2012 APR 1 2 2013

Figure B.3: IRB informed consent form (no payment)

Appendix C

Survey Questions

All of the survey questions used in the user study are listed here for convenience.

Question	Response Type
“Which control method was easier to learn?”	“Leader”, “Predator”, “Teammates” or “None”
“Which control method was easier to use?”	“Leader”, “Predator”, “Teammates” or “None”
“Which control method was most effective at getting the robots to the barrels?”	“Leader”, “Predator”, “Teammates” or “None”
“Which control method did you prefer?”	“Leader”, “Predator”, “Teammates” or “None”
“How many robots did you prefer to have in the group?”	“Many”, “Moderate” or “Few”
“Which number of robots was easiest to control?’ ’	“Many”, “Moderate” or “Few”
“Which number of robots was most effective?”	“Many”, “Moderate” or “Few”
“If you had to run another scenario, which control method would you want to use?”	“Leader”, “Predator”, or “Teammates”
“If you had to run another scenario, how many robots would you want to use?”	“Many”, “Moderate” or “Few”
“What strategies (if any) did you develop to control the robots with the Predator (fish)’ ?’	Free-response
“What strategies (if any) did you develop to control the robots with the Leader (plane)?”	Free-response
“What strategies (if any) did you develop to control the robots with Teammates (members of the group)?”	Free-response
“What strategies (if any) did you develop for controlling a large group of robots?”	Free-response
“What strategies (if any) did you develop for controlling a small group of robots?”	Free-response
“Any other comments or thoughts?”	Free-response

Table C.1: Post-experiment survey questions

Question	Response Type
"Sex"	"Male" or "Female"
"Age"	Numeric
"Do you have normal vision or corrected-to-normal vision?"	"Normal" or "Corrected to Normal"
"Are you colorblind?"	Yes/No
"'What is your level of experience working or playing with robots"	"extremely experienced" to "not at all experienced" (5 values)
"What is your level of experience playing video games"	"extremely experienced" to "not at all experienced" (5 values)

Table C.2: Pre-experiment demographic survey

Question	Response Type
"I was able to control the agents easily"	"strongly agree" to "strongly disagree" (7 values)
"The agents went where I wanted them to"	"strongly agree" to "strongly disagree" (7 values)
"This scenario was frustrating"	"strongly agree" to "strongly disagree" (7 values)
"This scenario was difficult"	"strongly agree" to "strongly disagree" (7 values)

Table C.3: Post-scenario survey

Question	Response Type
"How many times did you hear the cow say "Moo"?"	Numeric
"How confident are you in your answer?"	"extremely confident" to "not at all confident" (5 values)

Table C.4: Auditory secondary task survey

Appendix D

Statistical Data

This appendix contains the raw statistical data for the analysis of the primary user study described in Chapter 5. Descriptions of the metrics in each table can be found in Section 4.3.5. See Section 5.3.1 for a full description of our analysis. All tables in this appendix are in the format *significance* or *mean* ± *standard_deviation (significance)*, where *mean* represents the mean or difference in means and *significance* is the statistical significance of the Tukey-Kramer adjusted p-value.

D.1 Fixed Effects Tests: $Pr > F$

This section lists the adjusted p-value for fixed effect tests on each group of metrics (see Section 4.3.5). Table D.1 lists the effects we tested.

Effect	Type	Section
Scale	Independent variable	4.2.1
Style	Independent variable	4.2.1
Sex	Demographic	4.3.6
Vision	Demographic	4.3.6
Robot Exp	Demographic	4.3.6
Games Exp	Demographic	4.3.6
Sim Order	Learning effect	4.2.7

Table D.1: Fixed effect tests

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Controllability	0.1550	$< .0001$	0.7687	0.1592	0.1618	0.6439	0.0900
Frustration	0.0075	$< .0001$	0.7051	0.1726	0.8571	0.4196	0.2540
Difficulty	0.0126	$< .0001$	0.5892	0.2271	0.3215	0.5475	0.0766
Ease of control	0.2153	$< .0001$	0.4800	0.2216	0.4069	0.2523	0.7378

Table D.2: Fixed Effects Tests: $Pr > F$: Post-Scenario Surveys

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Information Gathered	$< .0001$	0.0002	0.6078	0.0085	0.0356	0.0212	0.4298
Barrels Depleted	$< .0001$	$< .0001$	0.4840	0.0583	0.0937	0.0829	0.6665
Score	$< .0001$	$< .0001$	0.5140	0.1306	0.1923	0.0944	0.5211

Table D.3: Fixed Effects Tests: $Pr > F$: Performance Metrics

Metric	Scale	Style
Number of Visual Events	0.4677	0.3557
Experiment Ordering	0.5384	0.0719
Framerate Hiccups	$< .0001$	0.4656
Number of Audio Events	0.3390	0.9972
Audio p(False Alarm)	0.1986	0.8837
Audio True Positives	0.2897	0.7067
Audio False Positives	0.4477	0.7772

Table D.4: Fixed Effects Tests: $Pr > F$: Possible Confounding Factors

Metric	Scale	Style	Sex	Vision	Robot	Games	Order
Total Mouse Movement	0.3934	$< .0001$	0.3555	0.0796	0.1836	0.9317	0.4031
Audio True Positives	0.2897	0.7067					
Audio Guess	0.2783	0.5195	0.4895	0.0924	0.3137	0.4057	0.4213
Audio Guess Error	0.9836	0.5891	0.0118	0.5916	0.0279	0.6197	0.0971
Audio Guess Error (absolute)	0.3008	0.0336	0.2572	0.8301	0.7109	0.7730	0.0347
Audio Guess Ratio	0.9464	0.5673	0.0113	0.2999	0.0094	0.6636	0.0547
Audio Percent Error	0.1784	0.0331	0.3685	0.6043	0.5806	0.6325	0.0355
Audio Confidence	0.4684	0.0009	0.6270	0.6170	0.3272	0.5046	0.0544
Visual False Positives	0.3192	0.5173	0.1158	0.5382	0.4833	0.0971	0.1538
Visual True Positives	0.4475	$< .0001$	0.6614	0.6197	0.7104	0.3518	0.0029
Visual Missed Detections	0.2338	0.0001	0.6664	0.6371	0.5867	0.3581	0.0022
Visual Response Time (no MD)	0.0821	$< .0001$	0.7379	0.6105	0.6596	0.5936	0.0739
Visual Response Time	0.1778	$< .0001$	0.9092	0.5976	0.5745	0.4525	0.0515

Table D.5: Fixed Effects Tests: $Pr > F$: Workload Metrics

Metric	Scale	Style	Sex	Vision	Robot	Games	Order
Total Neglect Time	0.6593	$< .0001$	0.5092	0.8823	0.2211	0.4226	0.9015
Longest Neglect Interval	0.0665	$< .0001$	0.8333	0.5404	0.1169	0.6245	0.8570
Average Neglect Time per Interval	0.2529	$< .0001$	0.5612	0.9447	0.2233	0.5996	0.0419
Number of Neglect Intervals	0.0070	0.0105	0.0147	0.6164	0.4226	0.2911	0.0110

Table D.6: Fixed Effects Tests: $Pr > F$: Neglect Time

Metric	Scale	Style	Sex	Vision	Robot Exp	Games Exp	Sim Order
Temporal Demand	0.4138	0.0011	0.4656	0.8135	0.5275	0.9372	0.5442
Effort	0.4964	0.0164	0.9819	0.5131	0.5682	0.4075	0.3938
Physical Demand	0.9443	0.0092	0.0719	0.1316	0.1775	0.8240	0.7123
Frustration	0.0937	$< .0001$	0.7191	0.7918	0.6520	0.3820	0.8882
Performance	0.1058	0.0010	0.2669	0.8341	0.0288	0.7014	0.1960
Mental Demand	0.7601	$< .0001$	0.8736	0.5972	0.4711	0.3382	0.3749
Overall	0.0828	0.1265	0.1377	0.3320	0.4004	0.3441	0.0013
Overall	0.1133	$< .0001$	0.0714	0.9030	0.4216	0.0339	0.3105
Overall (Raw Score)	0.1177	$< .0001$	0.6595	0.9877	0.5508	0.6475	0.8070

Table D.7: Fixed Effects Tests: $Pr > F$: NASA-TLX

D.2 Pairwise Comparisons

This section lists the differences in mean and associated significance values for each pairwise comparison across scale and control style. Descriptions of the metrics in each table can be found in Section 4.3.5. All tables in this section are in the format *difference_in_mean* ± *standard_deviation (significance)*.

D.2.1 Scale

	Scale		
Metric	20 vs 50	20 vs 100	50 vs 100
Controllability	-0.38 ± 0.20 (0.1470)	-0.10 ± 0.20 (0.8619)	0.28 ± 0.20 (0.3570)
Frustration	0.56 ± 0.17 (0.0052)	0.27 ± 0.17 (0.2650)	-0.29 ± 0.17 (0.2070)
Difficulty	0.51 ± 0.17 (0.0113)	0.14 ± 0.17 (0.6765)	-0.37 ± 0.17 (0.0912)
Ease of control	-0.35 ± 0.20 (0.1880)	-0.19 ± 0.20 (0.5976)	0.16 ± 0.20 (0.7014)

Table D.8: Pairwise Comparisons (Scale): Post-Scenario Surveys

	Scale		
Metric	20 vs 50	20 vs 100	50 vs 100
Information Gathered	-72.86 ± 12.07 ($< .0001$)	-96.47 ± 12.00 ($< .0001$)	-23.61 ± 12.13 (0.1366)
Barrels Depleted	-2.62 ± 0.29 ($< .0001$)	-3.05 ± 0.29 ($< .0001$)	-0.43 ± 0.29 (0.3209)
Score	-80.46 ± 12.16 ($< .0001$)	-101.74 ± 12.08 ($< .0001$)	-21.27 ± 12.22 (0.2003)

Table D.9: Pairwise Comparisons (Scale): Performance Metrics

Metric	Scale 20 vs 50	20 vs 100	50 vs 100
Number of Visual Events	0.35 ± 0.37 (0.6235)	0.44 ± 0.37 (0.4718)	0.09 ± 0.37 (0.9669)
Experiment Ordering	0.00 ± 0.00 ()	0.00 ± 0.00 ()	0.00 ± 0.00 ()
Framerate Hiccups	-1.50 ± 0.23 ($< .0001$)	-1.94 ± 0.23 ($< .0001$)	-0.44 ± 0.23 (0.1390)
Number of Audio Events	0.23 ± 0.17 (0.3568)	0.04 ± 0.17 (0.9733)	-0.19 ± 0.17 (0.4810)
Audio p(False Alarm)	-0.01 ± 0.01 (0.4782)	0.01 ± 0.01 (0.8019)	0.02 ± 0.01 (0.1795)
Audio True Positives	0.29 ± 0.42 (0.7673)	-0.37 ± 0.42 (0.6465)	-0.67 ± 0.42 (0.2595)
Audio False Positives	-0.07 ± 0.41 (0.9836)	0.42 ± 0.41 (0.5741)	0.49 ± 0.41 (0.4686)

Table D.10: Pairwise Comparisons (Scale): Possible Confounding Factors

Metric	Scale 20 vs 50	20 vs 100	50 vs 100
Total Mouse Movement	1006.73 ± 2219.21 (0.8930)	-2002.92 ± 2217.41 (0.6408)	-3009.65 ± 2222.98 (0.3727)
Audio True Positives	0.29 ± 0.42 (0.7673)	-0.37 ± 0.42 (0.6465)	-0.67 ± 0.42 (0.2595)
Audio Guess	0.33 ± 0.42 (0.7210)	-0.36 ± 0.42 (0.6695)	-0.69 ± 0.43 (0.2471)
Audio Guess Error	-0.05 ± 0.31 (0.9834)	-0.04 ± 0.31 (0.9904)	0.01 ± 0.31 (0.9990)
Audio Guess Error (absolute)	0.37 ± 0.23 (0.2722)	0.15 ± 0.23 (0.8036)	-0.22 ± 0.24 (0.6259)
Audio Guess Ratio	0.01 ± 0.04 (0.9569)	0.01 ± 0.04 (0.9544)	0.00 ± 0.04 (1.0000)
Audio Percent Error	0.05 ± 0.03 (0.1718)	0.04 ± 0.03 (0.3890)	-0.01 ± 0.03 (0.8693)
Audio Confidence	-0.11 ± 0.13 (0.6841)	0.05 ± 0.13 (0.9202)	0.16 ± 0.13 (0.4498)
Visual False Positives	1.51 ± 2.32 (0.7923)	-2.04 ± 2.32 (0.6548)	-3.56 ± 2.34 (0.2892)
Visual True Positives	-0.70 ± 0.61 (0.4832)	-0.06 ± 0.60 (0.9946)	0.64 ± 0.61 (0.5467)
Visual Missed Detections	0.85 ± 0.49 (0.2046)	0.39 ± 0.49 (0.7051)	-0.46 ± 0.50 (0.6247)
Visual Response Time (no MD)	0.00 ± 0.02 (0.9581)	-0.03 ± 0.02 (0.1658)	-0.03 ± 0.02 (0.0997)
Visual Response Time	0.03 ± 0.03 (0.4260)	-0.02 ± 0.03 (0.8146)	-0.05 ± 0.03 (0.1624)

Table D.11: Pairwise Comparisons (Scale): Workload Metrics

Metric	Scale		
	20 vs 50	20 vs 100	50 vs 100
Total Neglect Time	0.63 ± 1.90 (0.9420)	1.72 ± 1.89 (0.6388)	1.09 ± 1.91 (0.8355)
Longest Neglect Interval	0.84 ± 0.36 (0.0635)	0.61 ± 0.36 (0.2241)	-0.23 ± 0.37 (0.7979)
Average Neglect Time per Interval	0.04 ± 0.03 (0.5079)	-0.02 ± 0.03 (0.8484)	-0.05 ± 0.03 (0.2344)
Number of Neglect Intervals	-4.06 ± 2.72 (0.3025)	4.97 ± 2.71 (0.1676)	9.03 ± 2.73 (0.0049)

Table D.12: Pairwise Comparisons (Scale): Neglect Time

Metric	Scale		
	20 vs 50	20 vs 100	50 vs 100
Temporal Demand	2.90 ± 2.18 (0.3862)	1.75 ± 2.17 (0.7003)	-1.15 ± 2.19 (0.8606)
Effort	1.49 ± 2.15 (0.7678)	-1.07 ± 2.14 (0.8723)	-2.56 ± 2.16 (0.4665)
Physical Demand	0.01 ± 1.91 (1.0000)	-0.55 ± 1.90 (0.9542)	-0.57 ± 1.92 (0.9533)
Frustration	5.28 ± 2.37 (0.0764)	2.43 ± 2.36 (0.5610)	-2.85 ± 2.38 (0.4617)
Performance	5.45 ± 2.59 (0.0993)	3.83 ± 2.57 (0.3043)	-1.62 ± 2.60 (0.8098)
Mental Demand	1.40 ± 1.88 (0.7394)	0.67 ± 1.88 (0.9313)	-0.73 ± 1.89 (0.9225)
Overall	5.18 ± 2.56 (0.1175)	4.91 ± 2.55 (0.1415)	-0.27 ± 2.58 (0.9938)
Overall	3.11 ± 1.66 (0.1591)	-0.03 ± 1.69 (0.9999)	-3.14 ± 1.71 (0.1712)
Overall (Raw Score)	16.42 ± 7.79 (0.0983)	7.05 ± 7.75 (0.6370)	-9.37 ± 7.83 (0.4604)

Table D.13: Pairwise Comparisons (Scale): NASA-TLX

D.2.2 Control Style

Metric	Control Style L vs P	 L vs S	 P vs S
Controllability	1.91 ± 0.20	0.11 ± 0.20	-1.80 ± 0.20
	$(< .0001)$	(0.8511)	$(< .0001)$
Frustration	-1.50 ± 0.19	-0.10 ± 0.19	1.40 ± 0.19
	$(< .0001)$	(0.8595)	$(< .0001)$
Difficulty	-1.56 ± 0.19	-0.02 ± 0.19	1.54 ± 0.19
	$(< .0001)$	(0.9922)	$(< .0001)$
Ease of control	1.90 ± 0.21	-0.03 ± 0.21	-1.93 ± 0.21
	$(< .0001)$	(0.9877)	$(< .0001)$

Table D.14: Pairwise Comparisons (Control Style): Post-Scenario Surveys

Metric	Control Style L vs P	 L vs S	 P vs S
Information Gathered	55.19 ± 12.12	28.80 ± 12.22	-26.39 ± 12.04
	$(< .0001)$	(0.0572)	(0.0826)
Barrels Depleted	2.46 ± 0.30	0.40 ± 0.30	-2.06 ± 0.30
	$(< .0001)$	(0.3926)	$(< .0001)$
Score	75.67 ± 13.05	24.65 ± 13.13	-51.02 ± 12.98
	$(< .0001)$	(0.1561)	(0.0007)

Table D.15: Pairwise Comparisons (Control Style): Performance Metrics

Metric	Control Style L vs P	 L vs S	 P vs S
Number of Visual Events	0.05 ± 0.36 (0.9899)	-0.43 ± 0.36 (0.4676)	-0.48 ± 0.36 (0.3907)
Experiment Ordering	0.00 ± 0.00 ()	0.00 ± 0.00 ()	0.00 ± 0.00 ()
Framerate Hiccups	0.00 ± 0.11 (1.0000)	0.12 ± 0.11 (0.5314)	0.12 ± 0.11 (0.5314)
Number of Audio Events	-0.01 ± 0.21 (0.9978)	-0.01 ± 0.21 (0.9976)	-0.00 ± 0.21 (1.0000)
Audio p(False Alarm)	-0.00 ± 0.01 (1.0000)	0.00 ± 0.01 (0.9034)	0.00 ± 0.01 (0.9022)
Audio True Positives	-0.34 ± 0.40 (0.6837)	-0.19 ± 0.40 (0.8881)	0.15 ± 0.40 (0.9275)
Audio False Positives	0.29 ± 0.42 (0.7610)	0.18 ± 0.42 (0.9029)	-0.11 ± 0.42 (0.9591)

Table D.16: Pairwise Comparisons (Control Style): Possible Confounding Factors

Metric	Control Style L vs P	 L vs S	 P vs S
Total Mouse Movement	-8861.83 ± 1042.30 (< .0001)	1878.40 ± 1052.67 (0.1852)	10740.00 ± 1034.01 (< .0001)
Audio True Positives	-0.34 ± 0.40 (0.6837)	-0.19 ± 0.40 (0.8881)	0.15 ± 0.40 (0.9275)
Audio Guess	-0.46 ± 0.41 (0.5093)	-0.34 ± 0.41 (0.6949)	0.12 ± 0.41 (0.9531)
Audio Guess Error	0.27 ± 0.27 (0.5815)	0.20 ± 0.28 (0.7403)	-0.07 ± 0.27 (0.9653)
Audio Guess Error (absolute)	-0.59 ± 0.22 (0.0256)	-0.28 ± 0.22 (0.4239)	0.31 ± 0.22 (0.3325)
Audio Guess Ratio	0.04 ± 0.03 (0.5369)	0.02 ± 0.03 (0.8664)	-0.02 ± 0.03 (0.8426)
Audio Percent Error	-0.07 ± 0.03 (0.0251)	-0.04 ± 0.03 (0.3204)	0.03 ± 0.03 (0.4346)
Audio Confidence	0.41 ± 0.12 (0.0021)	0.03 ± 0.12 (0.9535)	-0.38 ± 0.11 (0.0048)
Visual False Positives	-3.74 ± 3.46 (0.5304)	-0.65 ± 3.47 (0.9807)	3.09 ± 3.45 (0.6471)
Visual True Positives	3.00 ± 0.88 (0.0038)	-1.16 ± 0.88 (0.3973)	-4.16 ± 0.88 (< .0001)
Visual Missed Detections	-2.92 ± 0.81 (0.0020)	0.55 ± 0.81 (0.7749)	3.48 ± 0.81 (0.0002)
Visual Response Time (no MD)	-0.07 ± 0.02 (0.0014)	0.03 ± 0.02 (0.3075)	0.10 ± 0.02 (< .0001)
Visual Response Time	-0.15 ± 0.04 (0.0003)	0.05 ± 0.04 (0.3554)	0.20 ± 0.04 (< .0001)

Table D.17: Pairwise Comparisons (Control Style): Workload Metrics

Metric	Control Style L vs P	L vs S	P vs S
Total Neglect Time	24.26 ± 2.08 (< .0001)	-14.44 ± 2.09 (< .0001)	-38.70 ± 2.08 (< .0001)
Longest Neglect Interval	2.21 ± 0.53 (0.0004)	-3.38 ± 0.54 (< .0001)	-5.60 ± 0.53 (< .0001)
Average Neglect Time per Interval	0.23 ± 0.07 (0.0066)	-0.28 ± 0.07 (0.0008)	-0.51 ± 0.07 (< .0001)
Number of Neglect Intervals	16.49 ± 5.24 (0.0077)	9.75 ± 5.25 (0.1618)	-6.74 ± 5.24 (0.4090)

Table D.18: Pairwise Comparisons (Control Style): Neglect Time

Metric	Control Style L vs P	L vs S	P vs S
Temporal Demand	-8.71 ± 2.32 (0.0013)	-2.02 ± 2.34 (0.6641)	6.69 ± 2.31 (0.0151)
Effort	-6.40 ± 2.76 (0.0625)	1.30 ± 2.77 (0.8865)	7.70 ± 2.76 (0.0200)
Physical Demand	-7.47 ± 3.18 (0.0580)	2.26 ± 3.18 (0.7581)	9.73 ± 3.17 (0.0096)
Frustration	-16.21 ± 2.17 (< .0001)	-2.11 ± 2.19 (0.6021)	14.09 ± 2.16 (< .0001)
Performance	-12.16 ± 3.20 (0.0011)	-2.75 ± 3.22 (0.6698)	9.41 ± 3.19 (0.0132)
Mental Demand	-10.85 ± 2.42 (0.0001)	1.41 ± 2.43 (0.8325)	12.25 ± 2.42 (< .0001)
Overall	-6.28 ± 3.26 (0.1409)	-1.03 ± 3.27 (0.9468)	5.25 ± 3.24 (0.2471)
Overall	-12.43 ± 2.12 (< .0001)	-0.86 ± 2.05 (0.9068)	11.56 ± 2.00 (< .0001)
Overall (Raw Score)	-61.17 ± 9.59 (< .0001)	-1.46 ± 9.62 (0.9874)	59.71 ± 9.56 (< .0001)

Table D.19: Pairwise Comparisons (Control Style): NASA-TLX

D.3 Means of User Study Metrics

This section lists the means and associated significance values for each value of scale and control style. Descriptions of the metrics in each table can be found in Section 4.3.5. All tables in this section are in the format *mean* ± *standard_deviation (significance)*.

D.3.1 Scale

Metric	20	50	100
Controllability	0.36 ± 0.29 (0.2114)	0.74 ± 0.29 (0.0125)	0.46 ± 0.29 (0.1105)
Frustration	1.07 ± 0.37 (0.0060)	0.51 ± 0.37 (0.1771)	0.80 ± 0.37 (0.0359)
Difficulty	1.27 ± 0.31 (0.0002)	0.76 ± 0.31 (0.0188)	1.13 ± 0.31 (0.0007)
Ease of control	0.03 ± 0.30 (0.9156)	0.38 ± 0.30 (0.2066)	0.22 ± 0.30 (0.4596)

Table D.20: Means of User Study Metrics (Scale): Post-Scenario Surveys

Metric	20	50	100
Information Gathered	317.41 ± 15.32 ($< .0001$)	390.27 ± 15.36 ($< .0001$)	413.88 ± 15.34 ($< .0001$)
Barrels Depleted	6.09 ± 0.34 ($< .0001$)	8.71 ± 0.34 ($< .0001$)	9.14 ± 0.34 ($< .0001$)
Score	322.45 ± 21.24 ($< .0001$)	402.92 ± 21.27 ($< .0001$)	424.19 ± 21.26 ($< .0001$)

Table D.21: Means of User Study Metrics (Scale): Peformance Metrics

Metric	20	50	100
Number of Visual Events	46.28 ± 0.26 (< .0001)	45.93 ± 0.26 (< .0001)	45.84 ± 0.26 (< .0001)
Experiment Ordering	5.05 ± 0.29 (< .0001)	4.75 ± 0.29 (< .0001)	5.20 ± 0.29 (< .0001)
Framerate Hiccups	0.50 ± 0.20 (0.0166)	2.01 ± 0.20 (< .0001)	2.44 ± 0.20 (< .0001)
Number of Audio Events	23.00 ± 0.13 (< .0001)	22.77 ± 0.13 (< .0001)	22.97 ± 0.13 (< .0001)
Audio p(False Alarm)	0.65 ± 0.01 (< .0001)	0.66 ± 0.01 (< .0001)	0.64 ± 0.01 (< .0001)
Audio True Positives	8.02 ± 0.30 (< .0001)	7.73 ± 0.30 (< .0001)	8.39 ± 0.30 (< .0001)
Audio False Positives	14.99 ± 0.30 (< .0001)	15.06 ± 0.30 (< .0001)	14.57 ± 0.30 (< .0001)

Table D.22: Means of User Study Metrics (Scale): Possible Confounding Factors

Metric	20	50	100
Total Mouse Movement	9088.66 ± 2986.45 (0.0037)	8081.92 ± 2987.83 (0.0093)	11092.00 ± 2987.38 (0.0005)
Audio True Positives	8.02 ± 0.30 (< .0001)	7.73 ± 0.30 (< .0001)	8.39 ± 0.30 (< .0001)
Audio Guess	8.25 ± 0.44 (< .0001)	7.92 ± 0.44 (< .0001)	8.61 ± 0.44 (< .0001)
Audio Guess Error	-0.21 ± 0.36 (0.5577)	-0.16 ± 0.36 (0.6609)	-0.17 ± 0.36 (0.6349)
Audio Guess Error (absolute)	1.10 ± 0.33 (0.0015)	0.74 ± 0.33 (0.0292)	0.96 ± 0.33 (0.0053)
Audio Guess Ratio	0.99 ± 0.04 (< .0001)	0.98 ± 0.04 (< .0001)	0.98 ± 0.04 (< .0001)
Audio Percent Error	0.15 ± 0.04 (0.0005)	0.10 ± 0.04 (0.0166)	0.11 ± 0.04 (0.0067)
Audio Confidence	2.72 ± 0.35 (< .0001)	2.83 ± 0.35 (< .0001)	2.67 ± 0.35 (< .0001)
Visual False Positives	0.59 ± 3.64 (0.8719)	-0.93 ± 3.64 (0.8004)	2.63 ± 3.64 (0.4734)
Visual True Positives	38.24 ± 2.38 (< .0001)	38.94 ± 2.38 (< .0001)	38.30 ± 2.38 (< .0001)
Visual Missed Detections	7.86 ± 2.27 (0.0011)	7.00 ± 2.27 (0.0034)	7.47 ± 2.27 (0.0019)
Visual Response Time (no MD)	0.81 ± 0.06 (< .0001)	0.81 ± 0.06 (< .0001)	0.84 ± 0.06 (< .0001)
Visual Response Time	1.04 ± 0.12 (< .0001)	1.01 ± 0.12 (< .0001)	1.06 ± 0.12 (< .0001)

Table D.23: Means of User Study Metrics (Scale): Workload Metrics

Metric	20	50	100
Total Neglect Time	48.37 ± 5.13	47.74 ± 5.13	46.65 ± 5.13
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Longest Neglect Interval	4.41 ± 0.90	3.57 ± 0.90	3.80 ± 0.90
	$(< .0001)$	(0.0002)	(0.0001)
Average Neglect Time per Interval	0.45 ± 0.12	0.42 ± 0.12	0.47 ± 0.12
	(0.0004)	(0.0009)	(0.0002)
Number of Neglect Intervals	109.00 ± 6.10	113.06 ± 6.10	104.02 ± 6.10
	$(< .0001)$	$(< .0001)$	$(< .0001)$

Table D.24: Means of User Study Metrics (Scale): Neglect Time

Metric	20	50	100
Temporal Demand	70.00 ± 7.25	67.10 ± 7.25	68.24 ± 7.25
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Effort	73.83 ± 6.69	72.34 ± 6.69	74.90 ± 6.69
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Physical Demand	42.23 ± 8.81	42.22 ± 8.81	42.78 ± 8.81
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Frustration	64.98 ± 6.56	59.70 ± 6.56	62.55 ± 6.56
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Performance	44.38 ± 6.72	38.93 ± 6.72	40.54 ± 6.72
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Mental Demand	79.86 ± 6.82	78.46 ± 6.82	79.18 ± 6.82
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Overall	57.24 ± 10.74	52.05 ± 10.74	52.33 ± 10.74
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Overall	66.89 ± 5.53	63.78 ± 5.57	66.92 ± 5.57
	$(< .0001)$	$(< .0001)$	$(< .0001)$
Overall (Raw Score)	375.17 ± 32.94	358.75 ± 32.94	368.12 ± 32.94
	$(< .0001)$	$(< .0001)$	$(< .0001)$

Table D.25: Means of User Study Metrics (Scale): NASA-TLX

D.3.2 Control Style

Metric	Leader	Predator	Stakeholders
Controllability	1.19 ± 0.29	-0.72 ± 0.29	1.09 ± 0.29
	(0.0001)	(0.0154)	(0.0004)
Frustration	0.26 ± 0.38	1.76 ± 0.38	0.36 ± 0.37
	(0.4900)	($< .0001$)	(0.3422)
Difficulty	0.53 ± 0.32	2.09 ± 0.32	0.55 ± 0.32
	(0.1028)	($< .0001$)	(0.0897)
Ease of control	0.83 ± 0.30	-1.06 ± 0.30	0.86 ± 0.30
	(0.0079)	(0.0009)	(0.0059)

Table D.26: Means of User Study Metrics (Control Style): Post-Scenario Surveys

Metric	Leader	Predator	Stakeholders
Information Gathered	401.85 ± 15.38	346.66 ± 15.33	373.05 ± 15.36
	($< .0001$)	($< .0001$)	($< .0001$)
Barrels Depleted	8.94 ± 0.34	6.47 ± 0.34	8.54 ± 0.34
	($< .0001$)	($< .0001$)	($< .0001$)
Score	416.63 ± 21.45	340.95 ± 21.42	391.98 ± 21.44
	($< .0001$)	($< .0001$)	($< .0001$)

Table D.27: Means of User Study Metrics (Control Style): Peformance Metrics

Metric	Leader	Predator	Stakeholders
Number of Visual Events	45.89 ± 0.26	45.84 ± 0.26	46.32 ± 0.26
	($< .0001$)	($< .0001$)	($< .0001$)
Experiment Ordering	5.35 ± 0.29	5.20 ± 0.29	4.46 ± 0.29
	($< .0001$)	($< .0001$)	($< .0001$)
Framerate Hiccups	1.69 ± 0.17	1.69 ± 0.17	1.57 ± 0.17
	($< .0001$)	($< .0001$)	($< .0001$)
Number of Audio Events	22.91 ± 0.15	22.92 ± 0.15	22.92 ± 0.15
	($< .0001$)	($< .0001$)	($< .0001$)
Audio p(False Alarm)	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01
	($< .0001$)	($< .0001$)	($< .0001$)
Audio True Positives	7.87 ± 0.30	8.21 ± 0.30	8.06 ± 0.30
	($< .0001$)	($< .0001$)	($< .0001$)
Audio False Positives	15.03 ± 0.30	14.73 ± 0.30	14.85 ± 0.30
	($< .0001$)	($< .0001$)	($< .0001$)

Table D.28: Means of User Study Metrics (Control Style): Possible Confounding Factors

Metric	Leader	Predator	Stakeholders
Total Mouse Movement	7092.91 ± 2765.82 (0.0134)	15955.00 ± 2763.47 ($< .0001$)	5214.50 ± 2764.78 (0.0651)
Audio True Positives	7.87 ± 0.30 ($< .0001$)	8.21 ± 0.30 ($< .0001$)	8.06 ± 0.30 ($< .0001$)
Audio Guess	7.99 ± 0.44 ($< .0001$)	8.45 ± 0.44 ($< .0001$)	8.33 ± 0.44 ($< .0001$)
Audio Guess Error	-0.02 ± 0.35 (0.9500)	-0.30 ± 0.35 (0.4044)	-0.23 ± 0.35 (0.5226)
Audio Guess Error (absolute)	0.65 ± 0.32 (0.0522)	1.23 ± 0.32 (0.0004)	0.92 ± 0.32 (0.0065)
Audio Guess Ratio	1.00 ± 0.04 ($< .0001$)	0.96 ± 0.04 ($< .0001$)	0.98 ± 0.04 ($< .0001$)
Audio Percent Error	0.08 ± 0.04 (0.0450)	0.15 ± 0.04 (0.0003)	0.12 ± 0.04 (0.0035)
Audio Confidence	2.89 ± 0.35 ($< .0001$)	2.47 ± 0.35 ($< .0001$)	2.85 ± 0.35 ($< .0001$)
Visual False Positives	-0.70 ± 3.93 (0.8596)	3.04 ± 3.93 (0.4427)	-0.05 ± 3.93 (0.9906)
Visual True Positives	39.11 ± 2.40 ($< .0001$)	36.11 ± 2.40 ($< .0001$)	40.26 ± 2.40 ($< .0001$)
Visual Missed Detections	6.65 ± 2.30 (0.0057)	9.58 ± 2.30 (0.0001)	6.10 ± 2.30 (0.0108)
Visual Response Time (no MD)	0.81 ± 0.06 ($< .0001$)	0.88 ± 0.06 ($< .0001$)	0.78 ± 0.06 ($< .0001$)
Visual Response Time	1.00 ± 0.12 ($< .0001$)	1.15 ± 0.12 ($< .0001$)	0.95 ± 0.12 ($< .0001$)

Table D.29: Means of User Study Metrics (Control Style): Workload Metrics

Metric	Leader	Predator	Stakeholders
Total Neglect Time	50.86 ± 5.16 ($< .0001$)	26.60 ± 5.16 ($< .0001$)	65.30 ± 5.16 ($< .0001$)
Longest Neglect Interval	3.54 ± 0.93 (0.0004)	1.32 ± 0.93 (0.1620)	6.92 ± 0.93 ($< .0001$)
Average Neglect Time per Interval	0.43 ± 0.12 (0.0011)	0.20 ± 0.12 (0.1081)	0.71 ± 0.12 ($< .0001$)
Number of Neglect Intervals	117.44 ± 6.63 ($< .0001$)	100.95 ± 6.62 ($< .0001$)	107.69 ± 6.63 ($< .0001$)

Table D.30: Means of User Study Metrics (Control Style): Neglect Time

Metric	Leader	Predator	Stakeholders
Temporal Demand	64.87 ± 7.27 (< .0001)	73.58 ± 7.27 (< .0001)	66.89 ± 7.27 (< .0001)
Effort	71.99 ± 6.77 (< .0001)	78.39 ± 6.77 (< .0001)	70.69 ± 6.77 (< .0001)
Physical Demand	40.67 ± 8.93 (< .0001)	48.14 ± 8.93 (< .0001)	38.41 ± 8.93 (< .0001)
Frustration	56.31 ± 6.54 (< .0001)	72.51 ± 6.54 (< .0001)	58.42 ± 6.54 (< .0001)
Performance	36.31 ± 6.81 (< .0001)	48.47 ± 6.81 (< .0001)	39.07 ± 6.81 (< .0001)
Mental Demand	76.02 ± 6.87 (< .0001)	86.86 ± 6.87 (< .0001)	74.61 ± 6.87 (< .0001)
Overall	51.44 ± 10.80 (< .0001)	57.72 ± 10.80 (< .0001)	52.46 ± 10.80 (< .0001)
Overall	61.44 ± 5.61 (< .0001)	73.86 ± 5.60 (< .0001)	62.30 ± 5.59 (< .0001)
Overall (Raw Score)	346.47 ± 33.10 (< .0001)	407.64 ± 33.10 (< .0001)	347.93 ± 33.10 (< .0001)

Table D.31: Means of User Study Metrics (Control Style): NASA-TLX

References

[1] J.A. Adams. Critical considerations for human-robot interface development. In *Proceedings of 2002 AAAI Fall Symposium*, pages 1–8. AAAI, 2002.

[2] L. Alboul, J. Saez-Pons, and J. Penders. Mixed human-robot team navigation in the GUARDIANS project. In *IEEE International Workshop on Safety, Security and Rescue Robotics*, pages 95–101. IEEE, 2008.

[3] B.E. Amrein. Human research and engineering directorate, major laboratory programs: Current thrust areas and recent research. Technical report, Army Research Lab Aberdeen Proving Ground MD, 2010. `http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA535097`.

[4] J. Artieda, J.M. Sebastian, P. Campoy, J.F. Correa, I.F. Mondragón, C. Martínez, and M. Olivares. Visual 3-D SLAM from UAVs. *Journal of Intelligent and Robotic Systems*, 55(4-5):299–321, 2009.

[5] L. Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, 1983.

[6] W.A. Bainbridge, J.W. Hart, E.S. Kim, and B. Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[7] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, 105(4):1232, 2008.

[8] L.E. Barnes, M.A. Fields, and K.P. Valavanis. Swarm formation control utilizing elliptical surfaces and limiting functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(6):1434–1445, 2009.

[9] S. Bashyal and G.K. Venayagamoorthy. Human swarm interaction for radiation source search and localization. In *IEEE Swarm Intelligence Symposium*, pages 1–8. IEEE, 2008.

[10] R.W. Beard and T.W. McLain. *Small unmanned aircraft: Theory and practice.* Princeton University Press, Princeton, New Jersey, 2012.

[11] E.R. Boer. Behavioral entropy as an index of workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 44, pages 125–128. SAGE Publications, 2000.

[12] O. Brafman and R.A. Beckstrom. *The starfish and the spider: The unstoppable power of leaderless organizations*. Penguin Group, New York, New York, 2006.

[13] A. Brown and Y. Lu. Performance test results of an integrated GPS/MEMS inertial navigation package. In *Proceedings of International Technical Meeting of the Satellite Division of the Institute of Navigation*, pages 825–832, 2004.

[14] B. Call, R.W. Beard, C.N. Taylor, and B. Barber. Obstacle avoidance for unmanned air vehicles using image feature tracking. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2006.

[15] L. Carlone, M.K. Ng, J. Du, B. Bona, and M. Indri. Rao-blackwellized particle filters multi-robot SLAM with unknown initial correspondences and limited communication. In *IEEE International Conference on Robotics and Automation*, pages 243–249. IEEE, 2010.

[16] J. G. Casali and W.W. Wierwille. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25(6):623–641, 1983.

[17] H. Chitsaz and S.M. LaValle. Time-optimal paths for a Dubins airplane. In *46th IEEE Conference on Decision and Control*, pages 2379–2384. IEEE, 2007.

[18] L. Conradt, J. Krause, I.D. Couzin, and T.J. Roper. "Leading according to need" in self-organizing groups. *The American Naturalist*, 173(3):304–312, 2009.

[19] N. Correll, J. Bachrach, D. Vickery, and D. Rus. Ad-hoc wireless network coverage with networked robots that cannot localize. In *IEEE International Conference on Robotics and Automation*, pages 3878–3885. IEEE, 2009.

[20] I.D. Couzin. Presentation at 2013 Science of Autonomy Workshop, Arlington, Virginia, April 2013. Obtained through correspondence.

[21] I.D. Couzin and J. Krause. Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, 32:1–75, 2003.

[22] I.D. Couzin, J. Krause, N.R. Franks, and S.A. Levin. Effective leadership and decision-making in animal groups on the move. *Nature: International Weekly Journal of Science*, 433(7025):513–516, 2005.

[23] I.D. Couzin, J. Krause, R. James, G.D. Ruxton, and N.R. Franks. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1):1–11, 2002.

[24] J.W. Crandall and M.L. Cummings. Developing performance metrics for the supervisory control of multiple robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 33–40. ACM, 2007.

[25] J.W. Crandall, C.W. Nielsen, and M.A. Goodrich. Towards predicting robot team performance. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 906–911. IEEE, 2003.

[26] C. Donald. Vigilance. In J. Noyes and M. Bransby, editors, *People in control: Human factors in control room design*, pages 35–50. The Institution of Engineering and Technology, London, United Kingdom, 2002.

[27] M.R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[28] M.A. Fields, E.C. Haas, S. Hill, C. Stachowiak, and L. Barnes. Effective robot team control methodologies for battlefield applications. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5862–5867. IEEE, 2009.

[29] S. Gan-Mor, R.L. Clark, and B.L. Upchurch. Implement lateral position accuracy under RTK-GPS tractor guidance. *Computers and Electronics in Agriculture*, 59(1):31–38, 2007.

[30] J. Gancet, E. Motard, A. Naghsh, C. Roast, M.M. Arancon, and L. Marques. User interfaces for human robot interactions with a swarm of robots in support to firefighters. In *IEEE International Conference on Robotics and Automation*, pages 2846–2851. IEEE, 2010.

[31] M.A. Goodrich, E.R. Boer, J.W. Crandall, R.W. Ricks, and M.L. Quigley. Behavioral entropy in human-robot interaction. Technical report, Proceedings of the Performance Metrics for Intelligent Systems Workshop, 2004. `http://www.dtic.mil/dtic/tr/fulltext/u2/a446467.pdf`.

[32] M.A. Goodrich, B. Pendleton, S. Kerman, and P.B. Sujit. What types of interactions do bio-inspired robot swarms and flocks afford a human. In *Proceedings of Robotics: Science and Systems*, 2012.

[33] M.A. Goodrich, B. Pendleton, P.B. Sujit, and J. Pinto. Toward human interaction with bio-inspired robot teams. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 2859–2864. IEEE, 2011.

[34] M.A. Goodrich, B. Pendleton, P.B. Sujit, J. Pinto, and J.W. Crandall. Toward human interaction with bio-inspired teams. In *The 10th International Conference on Autonomous Agents and Multiagent Systems*, volume 3, pages 1265–1266. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[35] M.A. Goodrich and A.C. Schultz. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.

[36] M.A. Goodrich, P.B. Sujit, S. Kerman, B. Pendleton, and J. Pinto. Enabling human interaction with bio-inspired robot teams: Topologies, leaders, predators, and stakeholders. BYU-HCMI technical report 2011-1. Technical report, Brigham Young University, 2011. `http://faculty.cs.byu.edu/~mike/mikeg/papers/CoherentTopologyProperties_v3.pdf`.

[37] Z.C. Gray. Communication free robot swarming. Master's thesis, Air Force Institute of Technology, 2009. `http://www.dtic.mil/dtic/tr/fulltext/u2/a500135.pdf`.

[38] E.C. Haas. Extreme scalability: Designing interfaces and algorithms for soldier-robotic swarm interaction. Technical report, Army Research Lab Aberdeen Proving Ground MD, 2009. `http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA523930`.

[39] E.C. Haas. Extreme scalability: Designing interfaces and algorithms for soldier-robotic swarm interaction, year 2. Technical report, Army Research Lab Aberdeen Proving Ground MD, 2010. `http://www.dtic.mil/dtic/tr/fulltext/u2/a498162.pdf`.

[40] S.G. Hart and L.E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, 1(3):139–183, 1988.

[41] G. Inalhan, F.D. Busse, and J.P. How. Precise formation flying control of multiple spacecraft using carrier-phase differential GPS. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2000.

[42] H. Jones and M. Snyder. Supervisory control of multiple robots based on a real-time strategy game interaction paradigm. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 1, pages 383–388. IEEE, 2001.

[43] S. Kerman. Methods and metrics for human interaction with bio-inspired swarms. Master's thesis, Brigham Young University, 2013. Under review for publication.

[44] S. Kerman, D. Brown, and M.A. Goodrich. Supporting human interaction with robust robot swarms. In *5th IEEE International Symposium on Resilient Control Systems*, pages 197–202. IEEE, 2012.

[45] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller. Multiple relative pose graphs for robust cooperative mapping. In *IEEE International Conference on Robotics and Automation*, pages 3185–3192. IEEE, 2010.

[46] Z. Kira and M.A. Potter. Exerting human control over decentralized robot swarms. In *4th International Conference on Autonomous Robots and Agents*, pages 566–571, 2010.

[47] T. Latif and A. Bozkurt. Line following terrestrial insect biobots. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 972–975. IEEE, 2012.

[48] S. Marras and M. Porfiri. Fish and robots swimming together: Attraction towards the robot demands biomimetic locomotion. *Journal of The Royal Society Interface*, 9(73):1856–1868, 2012.

[49] J.D. McLurkin. Stupid robot tricks: A behavior-based distributed algorithm library for programming swarms of robots. Master's thesis, Massachusetts Institute of Technology, 2004. `http://people.csail.mit.edu/jamesm/publications/McLurkin-SM-MIT-2004(72dpi).pdf`.

[50] J.D. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt. Speaking swarmish: Human-robot interface design for large swarms of autonomous mobile robots. In *AAAI Spring Symposium*, 2006.

[51] M. Mesbahi and M. Egerstedt. *Graph theoretic methods in multiagent networks*. Princeton University Press, Princeton, New Jersey, 2010.

[52] N. Meshkati, P.A. Hancock, M. Rahimi, and S.M. Dawes. Techniques in mental workload assessment. In J.R. Wilson and E.N. Corlett, editors, *Evaluation of human work: A*

practical ergonomics methodology, pages 749–782. CRC Press, Boca Raton, Florida, 2nd edition, 1995.

[53] N. Michael, M.M. Zavlanos, V. Kumar, and G.J. Pappas. Distributed multi-robot task assignment and formation control. In *IEEE International Conference on Robotics and Automation*, pages 128–133. IEEE, 2008.

[54] G.D. Ogden, J.M. Levine, and E.J. Eisner. Measurement of workload by secondary tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 21(5):529–548, 1979.

[55] R. Olfati-Saber. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on Automatic Control*, 51(3):401–420, 2006.

[56] D.R. Olsen and M.A. Goodrich. Metrics for evaluating human-robot interactions. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, volume 2003, page 4, 2003.

[57] D.R. Olsen and S.B. Wood. Fan-out: Measuring human control of multiple robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 231–238. ACM, 2004.

[58] B.W. Parkinson and P.K. Enge. Differential GPS. In B.W. Parkinson and J.J. Spiker, editors, *Global positioning system: Theory and applications*, volume 2, pages 3–50. AIAA, Reston, Virginia, 1996.

[59] J. Penders. GUARDIANS final report. Technical report, Sheffield Hallam University, 2010. `http://shura.shu.ac.uk/2340/1/FinalReportMarch.pdf`.

[60] W. Ren and R.W. Beard. *Distributed consensus in multi-vehicle cooperative control: theory and applications*. Springer, New York, New York, 2008.

[61] E. Şahin. Swarm robotics: From sources of inspiration to domains of application. In E. Şahin and W.M. Spears, editors, *Swarm Robotics*, pages 10–20. Springer, New York, New York, 2005.

[62] M.S. Sanders and E.J. McCormick. *Human factors in engineering and design*. Mcgraw-Hill Book Company, San Francisco, California, 7th edition, 1993.

[63] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The International Journal of Robotics Research*, 21(8):735–758, 2002.

[64] T. Sheridan. Human supervisory control of robot systems. In *IEEE International Conference on Robotics and Automation*, volume 3, pages 808–812. IEEE, 1986.

[65] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M.A. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pages 33–40. ACM, 2006.

[66] A. Steinfeld, O.C. Jenkins, and B. Scassellati. The Oz of Wizard: Simulating the human for interaction research. In *4th ACM/IEEE International Conference on Human-Robot Interaction*, pages 101–107. IEEE, 2009.

[67] D.J.T. Sumpter. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):5, 2006.

[68] D.J.T. Sumpter. *Collective animal behavior*. Princeton University Press, Princeton, New Jersey, 2010.

[69] C.D. Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177, 2002.

[70] Wikipedia. Self-separation (aircraft). `http://en.wikipedia.org/wiki/Self_Separation_(aircraft)`, Feb 2011.

[71] A.F.T. Winfield, C.J. Harper, and J. Nembrini. Towards dependable swarms and a new discipline of swarm engineering. In E. Şahin and W.M. Spears, editors, *Swarm Robotics*, pages 126–142. Springer, New York, New York, 2005.

[72] M.M. Zavlanos, H.G. Tanner, A. Jadbabaie, and G.J. Pappas. Hybrid control for connectivity preserving flocking. *IEEE Transactions on Automatic Control*, 54(12):2869–2875, 2009.