Contents lists available at ScienceDirect

# International Journal of Human - Computer Studies

# Developing Future Wearable Interfaces for Human-Drone Teams through a Virtual Drone Search Game

Marlena R. Fraune [*,a], Ahmed S. Khalaf [b,c,e], Mahlet Zemedie [b], Poom Pianpak [b],
Zahra NaminiMianji [b,c], Sultan A. Alharthi [b,c,f], Igor Dolgov [a], Bill Hamilton [b,d], Son Tran [b], Phoebe
O. Toups Dugas[b,c]

[a] *Department of Psychology, New Mexico State University, Las Cruces, New Mexico, USA*
[b] *Department of Computer Science, New Mexico State University, Las Cruces, New Mexico, USA*
[c] *Play & Interactive Experiences for Learning Lab, USA*
[d] *Participatory Live Experiences Lab, USA*
[e] *Department of Computer Science, Al Baha University, Saudi Arabia*
[f] *Department of Software Engineering, University of Jeddah, Jeddah, Saudi Arabia*

## ARTICLE INFO

## ABSTRACT

Autonomous robotic vehicles (i.e., drones) are potentially transformative for search and rescue (SAR). This paper works toward wearable interfaces, through which humans team with multiple drones. We introduce the Virtual Drone Search Game as a first step in creating a mixed reality simulation for humans to practice drone teaming and SAR techniques. Our goals are to (1) evaluate input modalities for the drones, derived from an iterative narrowing of the design space, (2) improve our mixed reality system for designing input modalities and training operators, and (3) collect data on how participants socially experience the virtual drones with which they work. In our study, 17 participants played the game with two input modalities (Gesture condition, Tap condition) in counterbalanced order. Results indicated that participants performed best with the Gesture condition. Participants found the multiple controls challenging, and future studies might include more training of the devices and game. Participants felt like a team with the drones and found them moderately agentic. In our future work, we will extend this testing to a more externally valid mixed reality game.

## 1. Introduction

As autonomous robotic vehicles (i.e., drones) proliferate, scholars expect them to be valuable to search and rescue (SAR[1]) Alharthi et al. (2018a); Khan and Neustaedter (2019) as they move from nuisance ABC News (2017); Hutson (2017) to valued teammates. As drones become more automated, we expect it to be useful for one person to direct multiple drones while maintaining situational awareness Endsley (1995). This pushes drones from being an extension of an individual pilot to members of a team and allows human operators to move from direct piloting via physically large, laptop-based user interface (UI) to wearable UIs that support mobility and situational awareness.

Our long-term objective is to build such wearable UIs for human-drone teams, predicting this need in the near future. To do so we build and test an intermediate step: a mixed reality that combines wearable UIs and humans operating in the physical world with a set of virtual drones. Such a *mobile laboratory* Ashbrook et al. (2009) is an ecologically valid research environment for developing algorithms for drones, UIs for directing drones, and hardware configurations that make up composite wearable computers (i.e., those assembled from heterogeneous pieces of hardware). This intermediate step enables us to design such systems without yet needing to consider the legal and safety ramifications of drone flight Dolgov and Hottman (2011); Dorr (2018); Hobbs (2010), which is outside our scope. The **objective of the present research** is to narrow down the space of needed devices to achieve such an interface by running a *fixed laboratory study* Ashbrook et al. (2009) using a 3D game UI, the *Virtual Drone Search Game*, to represent the physical world and considering wearable computer configurations. We develop the following research questions (RQs):

1. To improve technological mediation of wearable multiple-drone control, we contrast two input modality conditions. **Which input modality will work more effectively?** As an exploratory study, we do not make a directional hypothesis.
2. We aim to further improve our mixed reality system for designing future wearable drone UIs and training operators. **How can we improve our system to maximize knowledge gain from designs and to make the experience compelling and stressful (i.e., like real SAR scenarios) to participants?**
3. As drones become more automated, they transition from tool to teammate. **What is a baseline for how participants socially experience the (virtual) drones when working with them?**

To address the research questions, we compare input modality design. To narrow the design space, we performed grounded theory on hardware that could be composed to create a wearable computer system. Taking this analysis, along with our target domain, we ran a pilot study with three device configurations. Following the pilot, we identified two candidate input modality combinations that we believed would be effective:

- **Gesture condition:** Head-mounted display (HMD), wrist-worn touchscreen, Leap Motion free-air gesture controller, and Twiddler 3 chording keyboard; and
- **Tap condition:** HMD, wrist-worn touchscreen, Tap finger-worn keyboard, and handheld mouse.

In this paper, we test human interaction with our UI to team with multiple drones to understand the design of future systems. We compare workload, performance, and situational awareness between two UIs to demonstrate the usefulness of our simulation for user testing and to learn which UI is most effective. We use free response questions to determine how to improve the system and game before we move it to a mixed reality system. Further, we examine how participants perceive the drones as teammates and potentially anthropomorphic agents.

The present research contributes ways to advance designing wearables to support human-drone teams. We find that: (RQ1) the Gesture condition produced better performance and lower workload than the Tap condition. However, across conditions, situational awareness of the drones was low. (RQ2) Based on participant comments, certain aspects of the game were more confusing than others (e.g., drone charging) and could be improved. (RQ3) Finally, participants felt like a group or team with the virtual drones, and rated them as moderately agentic.

The remainder of the paper is organized as follows. Multiple background sections cover SAR, human-drone teaming, wearables, and game design. We describe our research artifact, the *Virtual Drone Search Game*, providing details on game design and implementation, including how technologies are assembled. Our methods section provides insight into the empirical user study, and the results section organizes results according to our research questions. We then provide a discussion and conclusion, pointing to future work and its value to SAR.

## 2. Research Context: SAR

*Search and rescue* (SAR) is a disaster response operation to locate persons who are in distress or imminent danger, aid them (e.g., medical, food), and move them to a safe place Department of Defense (2006). There are different types of SAR operations (e.g., urban search and rescue[2], mountain rescue[3]). Time is critical in SAR operations; using technology (e.g., drones) could decrease the time needed to find persons who are in distress Waharte and Trigoni (2010). Most studies on drones for SAR work with simulation Cacace et al. (2016, 2017); Karaca et al.

(2018), but a case study found that using a commercial drone to assist search for a missing climber near the peak of the Karakoram Mountains reduced the number of people needed to find the climber and reduced the risk of those SAR responders McRae et al. (2019). Conditions like this are especially dangerous to people because of the vast area needed to search and dangerous altitude. Findings that drones could also decrease the time needed to find missing persons Weldon and Hupy (2020) makes drones promising to the field of SAR (see Section 3.1 for more on teaming with drones).

Constant training is core to SAR Fischer et al. (2014); Toups et al. (2016); Toups and Kerne (2007). Traditional methods, such as high-fidelity simulations and classroom courses, may fall short of providing effective and cost-efficient training that is needed for emerging technologies. Advances in technology open new opportunities, including computer-based training simulations, drones, mixed reality, virtual reality (VR), and wearables. Prior studies report on the usefulness of these technologies for training purposes Alharthi et al. (2018b); Hsu et al. (2013); Toups et al. (2011). Hsu et al. Hsu et al. (2013) provide an overview of how VR has been used for disaster preparedness training. These types of simulations also provide training opportunities for human-agent coordination and collaboration Fischer et al. (2017); Ramchurn et al. (2016), helping responders to build advanced coordination skills. Advances in computers and wearable technologies have the potential to enhance the design of mixed-reality training Feese et al. (2013).

Our ongoing work develops a mixed reality game and wearable system to test safe use of drones across contexts to help train and prepare for SAR. Our ultimate goal is to examine how good training with this technology can improve real-world SAR. In this work, we first validate the concept by testing the game on a computer prototype.

## 3. Background: Human-Drone Teaming

The term *drone*, used for simplicity, refers to an aerial vehicle that can be controlled remotely Austin (2010); Barin et al. (2017); Chang et al. (2017); Jones et al. (2016). Drones have different flight capabilities that depend on size (e.g., palm-sized to small-jet-sized) and flight type (e.g., quadcopter, hexacopter, fixed-wing) Vergouw et al. (2016). Payloads are the equipment that drones carry to perform useful work (e.g., camera, thermal imager, GPS) Austin (2010).

Drones have high incident rates (e.g., crashes), approximately 10 times greater than their crewed counterparts Dolgov and Hottman (2011); Hobbs (2010). Although equipment malfunctions and weather contribute, the primary cause is pilot error Chao et al. (2010); Dolgov and Hottman (2011); Hobbs (2010). Poor pilot performance often stems from lack of situational awareness Endsley (1995, 2000) and / or high workload Dolgov et al. (2017). As with most UIs, while pilots generally bear the blame, often suboptimal control interface design and / or training are the true source of error Norman (2013); Pancock et al. (2016). The system we are testing will allow researchers and practitioners safe testing of UIs before using a system in the real world.

Drones are expected to play a crucial role in SAR and are already involved in the field. Drones were, controversially used for SAR during Hurricane Harvey in summer 2017 ABC News (2017); Hutson (2017) to create a 3D map of the flooding and damage.

In the USA, one person can legally pilot only one drone at a time Dorr (2018); however, such limits may reduce SAR efficiency.

In this paper, we test a computer prototype of the system for a person to control multiple drones at once, forming a human-drone team with a single human operator. Once researchers demonstrate sufficiently low incidence rates, it may be beneficial to increase the legal limit of number of drones per pilot to increase efficiency of SAR teams.

### 3.1. Multi-Drone Interaction

Existing work has primarily focused on interactions with a single

---

[2] FEMA USAR: https://www.fema.gov/urban-search-rescue
[3] Mountain Rescue Association: http://mra.org

ARV; interacting with a *team* of ARVs is just being explored Brambilla et al. (2013); Chen and Barnes (2012); Chien et al. (2010); Kira and Potter (2009); Kolling et al. (2012); Penders et al. (2011). Research on teams has focused on the technical side of how ARVs might work together Modares et al. (2017); Remes et al. (2013); Sanchez-Lopez et al. (2014), with little research on how an individual human can control multiple ARVs. What research there is on human integration has begun to explore how to appropriately signal operators when multi-ARV teams require attention Chien et al. (2011) and the use of multitouch and gestures to guide operation Hayes et al. (2010); Micire et al. (2009). In one system, a group of two to three ARVs fly autonomously while a human focuses mostly on SAR Bevacqua et al. (2015); Cacace et al. (2016, 2017). That human is given the ability to take over control of ARVs at a low (e.g., "move a little more this direction") or high (e.g., "certified area") level. Individuals interact using a mixed-initiative multimodal system Cacace et al. (2016).

Systems that rely on drone autonomy have some challenges for current technology. They assume that the SAR expert is also familiar with the drone-controlling system Cacace et al. (2017). However, if teams must quickly deploy multiple drones in emergency situations, the responders may have little to no training, or may have been trained a long time ago.

In this project, our goal is to develop a system that requires less training for first-time users and that requires less computer autonomy for directing the drones. This project differs from swarm control because the pilot can control multiple drones independent of each other.

### 3.2. Anthropomorphic Perceptions of Drones and the Human-Drone Team

In human-robot interaction, social responses have been critical for human-robot teamwork and defining human expectations of robots Fraune et al. (2017a); Lee et al. (2010). While we do not expect to find differences in these measures across conditions, we use them as a baseline for future studies in which participants may perceive robots as differently anthropomorphic, negative or positive, and like teammates.

#### 3.2.1. Anthropomorphism

*Anthropomorphism* refers to how humans perceive robots to be similar to humans (e.g., in appearance, in behavior) Epley and Waytz (2010); Epley et al. (2007). The more human-like people perceive robots to be, the greater humans' expectations Lee et al. (2010). For example, people expect a robot that looks or behaves like a human to be able to engage in conversation, even if the robot cannot. Certain types of anthropomorphism occur with robotic teammates even if the robots do not look like humans; e.g., people have held funerals for mechanomorphic (i.e., machine-like) military robots Carpenter (2016), and dog-like robots Robertson (2018). Given the effects of anthropomorphism on human behavior, we measure social responses to robots.

Anthropomorphism is divided into *experience* (i.e., the ability to feel positive or negative stimuli like pleasure and pain) and *agency* (i.e., the ability to plan, calculate, or act) Haslam et al. (2008); Kozak et al. (2006). Robots are often viewed as low in experience Gray et al. (2007); Wullenkord et al. (2016) but medium-to-high agency Kahn Jr. et al. (2012); Lee and Lau (2011). As perceptions of an agent's experience increase, so do humans' desire to act in a humane way toward the agent Haslam et al. (2008); Waytz et al. (2010). High perceptions of robot experience relate to putting more effort into keeping them safe, which may alter how much danger people allow SAR drones to get into. Perceptions of an agent's agency relate to how much people expect an agent to be able to act on its own Haslam et al. (2008); Waytz et al. (2010). Higher perceptions of agency should relate to lower amounts of human oversight. Because in this paper, participants direct the drones, it is likely that they will report low levels of perceived drone experience and agency. Measurements in this study can serve as a baseline for future studies in which participants have less control over the drones.

#### 3.2.2. Entitativity

Campbell describes *entitativity* as a group being cohesive, or like a single entity Campbell (1958). Entitativity can be applied to robots Fraune et al. (2017b) and even "alien" shapes Dasgupta et al. (1999). High-entitativity groups have members that are typically have high similarity in appearance, common goals, and interdependence or shared outcomes. High group entitativity relates to team member consistency, meaning that group members are more likely to interact with each other in the future Castano et al. (2003b); Haslam et al. (2013); Insko et al. (2013); Lickel et al. (2001a). It also relates to positive behavior toward group members, whether they are humans Castano et al. (2003b); Haslam et al. (2013); Lickel et al. (2001a) or robots Fraune et al. (2017b). In long-term interaction or as experiment stakes increase in future studies, we expect people to feel more entitative with robots.

#### 3.2.3. Emotion

*Emotion* relates to behavior; people tend to approach things to which they have positive emotions and avoid things to which they have negative or anxious emotions Cottrell and Neuberg (2005a). *Group-based emotion* is felt with others (e.g., as an individual, one might have a negative emotion toward football but, as a group, feel proud of a local team) Smith and Mackie (2008); Smith et al. (2007).

Based on theories of emotions and group-based emotions, participants would be more likely to continue interacting with drones that they have more positive and fewer negative emotions about. Research on human interaction with technology has found that people with anxiety or fear related to technology, rated on scales like the Computer Anxiety Scale (CAS) Marcoulides (1989), the Computer Anxiety Rating Scale (CARS) Heinssen Jr. et al. (1987), and the Robot Anxiety Scale (RAS) Nomura et al. (2006) predicts have lower performance dos Santos and Santana (2018), satisfaction, and intention to use the technology later Meuter et al. (2003). Further, positive emotions toward technology predicts willingness to interact with it even more than negative emotions does Smith et al. (2020). To present a full view of people's emotions underlying their interactions with the drones in this study, we measured both positive and negative emotions. Emotions toward using drones is important as drones become increasingly useful in SAR situations.

## 4. Background: Wearable Computers

*Wearable computers* are equipped on various locations on a person's body Barfield (2015); Mann (1997); Starner et al. (1997). These devices establish constant interaction between the environment and the user and often form their own network of intercommunicating effectors and sensors. Wearable input devices vary widely in terms of how they acquire input from a user. Some systems include mini-QWERTY keyboards or virtual keyboards and pointing devices, mimicking desktop designs, but wearables open up a range of possibilities for full-body interaction and sensor-based, and context-aware designs Bellotti and Edwards (2001). For output, a number of displays, haptics, and audio feedback UIs exist. A key concern centers on how components inhibit mobility.

Wearable devices have been minimally used in disaster response (e. g., firefighters may wear a TRX System to track individuals where GPS fails Systems (2016)).

Using drones and wearables together may provide great benefits to first responders. For example, Epson's MOVERIO Augmented Reality Smart Glasses have been used in simulated disaster situations Epson (2016). In their experiment, they use a drone to capture visuals of the disaster area, which are sent to the rescue team in the field, who use the smart glasses, to help with decision making.

The present research aims at creating testbeds for *composite* wearable computers in the context of working with drones. By *composite wearable computers*, we mean assembling devices, primarily wearables, but also devices that can be repurposed as such, for a particular application. For a number of future applications, it is useful to consider how wearable can

be formed as a composite of effectors and sensors that work in harmony, but are modular, providing support to the wearer's primary task. Because the present research is aimed at future systems, it is important for us to consider the range of potential wearable systems, rather than existing, potentially restrictive platforms (e.g., Google Glass, Apple Watch). In this research, we tested different wearable devices against each other. See the Wearable Interface: Designing from the Framework Section for details.

## 5. Background: Game Design Terminology

Games are framed as a combination of rules and play, involving designed game mechanics, through which players make choices Salen and Zimmerman (2004). *Rules* are the structures of a game that constrain player choices, while *play* is the freedom to make choices within those constraints Salen and Zimmerman (2004). Rules define the outcomes of choices, resulting in new, observable game states. To that end, play is the essential experience of the system that the rules create. The combination of rules and play leads to designed moments of choice for players: *game mechanics* Adams and Dormans (2012); Juul (2005); Salen and Zimmerman (2004).

Jørgensen Jørgensen (2013) defines "gameworld" as "an information space and an ecological environment designed with certain gameplay activities in mind" (Jørgensen, 2013, p23–24). They are thus virtual spaces, inhabited by avatars, that serve as UIs to a game system. The gameworld enforces the rules of the game.

The present research uses a virtual gameworld, but aims at building a mixed reality game. In the mixed reality, the gameworld is created by augmenting the physical world. The player experiences reality first-person, but experiences the gameworld through the wearable UI. The current version of this research works with a first-person perspective on a gameworld that is reflective of a location in the physical world.

## 6. Research Artifact: Virtual Drone Search Game

The present research designs for future SAR scenarios, in which human-drone teams are essential. We expect such scenarios to involve human operators moving through hazardous environments and relying on drones for remote intelligence about where to search and where hazards are located. To that end, we develop a *Virtual Drone Search Game* in which the player, directing a team of virtual drones, must identify hidden locations in a physical-world environment while avoiding hazards. We build the system as a game, aiming to make the experience enjoyable to participants and to create stress Salen and Zimmerman (2004); Toups et al. (2011).

Creating stress is important because search and rescue (SAR) situations can be high-stress, and people respond differently under high- than low-stress Westman and Eden (1996). Prior studies designing games to support disaster response aimed to create stressful environments to match real-world experience. Therefore, in our simulation, participants experiencing some stress advances ecological validity of the design.

For this study, we use a desktop apparatus that simulates wearable interfaces by incorporating some of the wearable hardware. Using the desktop apparatus simplifies deployment and creates less burden for participants at the cost of ecological validity. The study gives us valuable insight to design wearables and enables us to ensure that the mixed reality game is as effective as possible. In the desktop apparatus, the player uses wearable hardware to get information about the drones and to provide direction to them. The player uses a laptop display as a first-person UI onto a gameworld that functions as an analog to the physical world.

While the present game apparatus is configured as a desktop application, we developed the prototype system such that it can be reconfigured as a mixed reality game for future studies. Thus, the system is complex, connecting multiple software and hardware systems; Fig. 1 shows how components communicate. To provide game logic and a first-

person gameworld UI through an avatar, we use an engine built on Unity (version 2017.3.1f1[4]). A drone simulation platform, Gazebo (version 7.0.0[5]), tracks and simulates the virtual drones. A custom planner provides the drones' intelligence, programmed using a Planning Domain Description Language (PDDL) McDermott et al. (1998), and communicates with Gazebo using the Robot Operating System (ROS, version Kinetic Kame[6]). Finally, a set of hardware and software interfaces connect with these components to create a UI that provides information about the gameworld and the ability to interact with the virtual drones. We use NASA WorldWind[7] for a map visualization, displayed on a wrist-worn touchscreen (Fig. 2), and provide drone data on an HMD (Fig. 4).

### 6.1. Game Objective

The objective of the game is to find hidden goal objects within multiple structures of a physical built environment. We derive this objective from SAR practice Fischer et al. (2015); Toups et al. (2016). All objects can be found by the drones, but some can only be collected by the drone and some by the player. The purpose of this design is twofold: (1) It represents reconnaissance drones' expected ability to locate victims and the need for SAR responders to render most aid and (2) It sets up the game mechanics to focus the player on working with the drones. The design begins to address SAR methods, with the player needing to consider trade-offs between performing activities themselves or having a drone perform the activity.

Each collected goal object scores points for the player. We designed the game so that there are more goals than we expect the players to be able to collect; this enables players to compete for a high score (Table 3) and ensures that the game sessions do not end early, maximizing data collection.

We based the play environment on a map of a physical-world location (for the mixed reality version, each structure exists physically in the game setting). Structures may be empty, contain a goal that the drone can locate (but only the player can collect), or a goal that the drone can locate and collect. The player needs to search each structure, either by moving their avatar to the structure or by sending a drone, to find its contents. Players or drones must then collect the goals that they are capable of collecting.

### 6.2. The Game

In the *Virtual Drone Search Game*, players identify structures in the game world that contain goals. Players need to move and use drones that function semi-autonomously to efficiently and effectively search the area.

More than one drone can move at a time. The drones remain still until participants directed them to complete tasks. When users issue a command, they need to select a drone to carry out the command. Participants make these five commands and send them to the server (Table 1). The drones don't get too close to each other to avoid a collision. Below, we describe the game further, with embedded explanations about how the future mixed reality version will differ. There are six actions used by ROS to execute the generated plan (Table 2).

#### 6.2.1. Rules and Game Mechanics

Players move their avatar and direct their drones to find and collect goal objects. Time is the main constraint on actions: the game concludes when time runs out, limiting the number of structures the player can visit and driving a need to efficiently use the drones. Because pausing to direct the drones will slow down the player, they are encouraged to
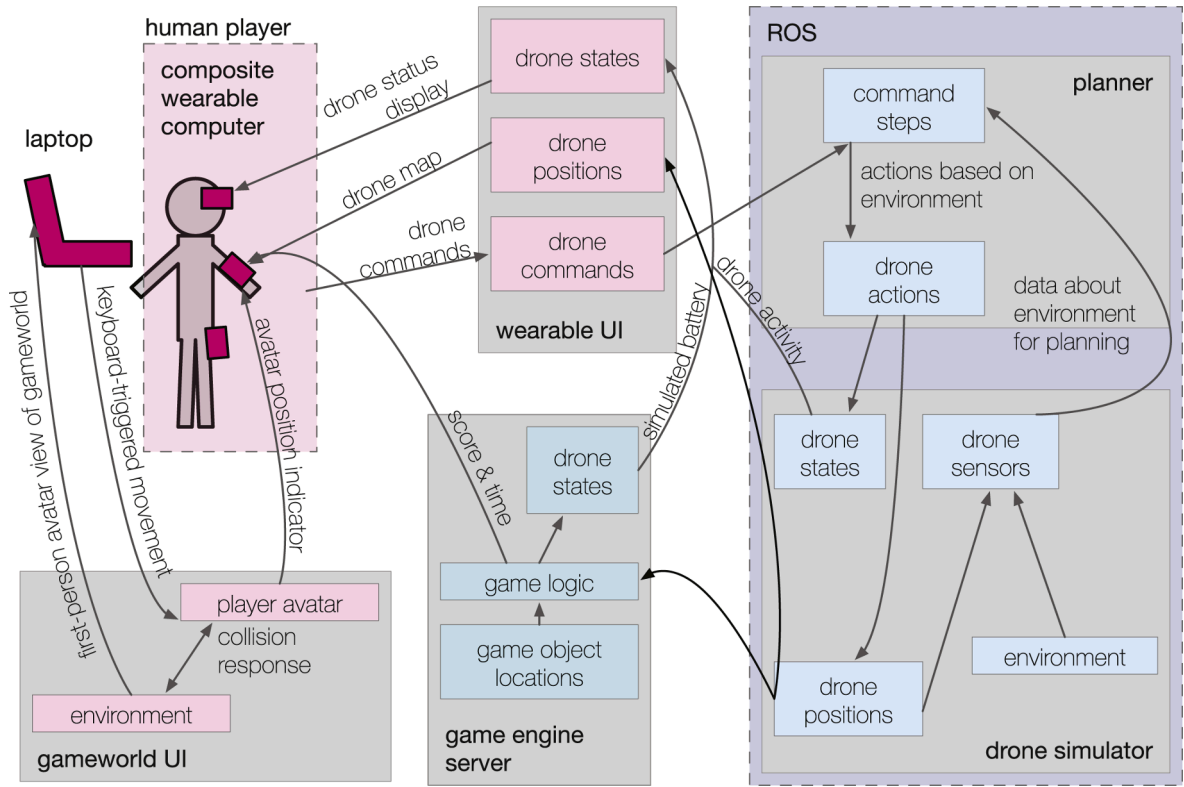
---

**Fig. 1.** Block diagram of the system architecture; gray blocks indicate separate components with interacting processing and data elements within. Multiple components run on multiple computers, communicating over network connections. Communication is indicated with directed arrows; non-obvious communication streams have additional labels. The player uses the composite wearable and the laptop to play the game, and their software elements are broken down in pink in the left portions of the diagram. A game engine server determines what goal objects are found and how they are scored, in green, near the middle. The drone simulator and planner (blue, right side) communicate with each other using ROS and collectively manage interpreting player commands and moving drones in a way that is constrained by the Gazebo simulator.
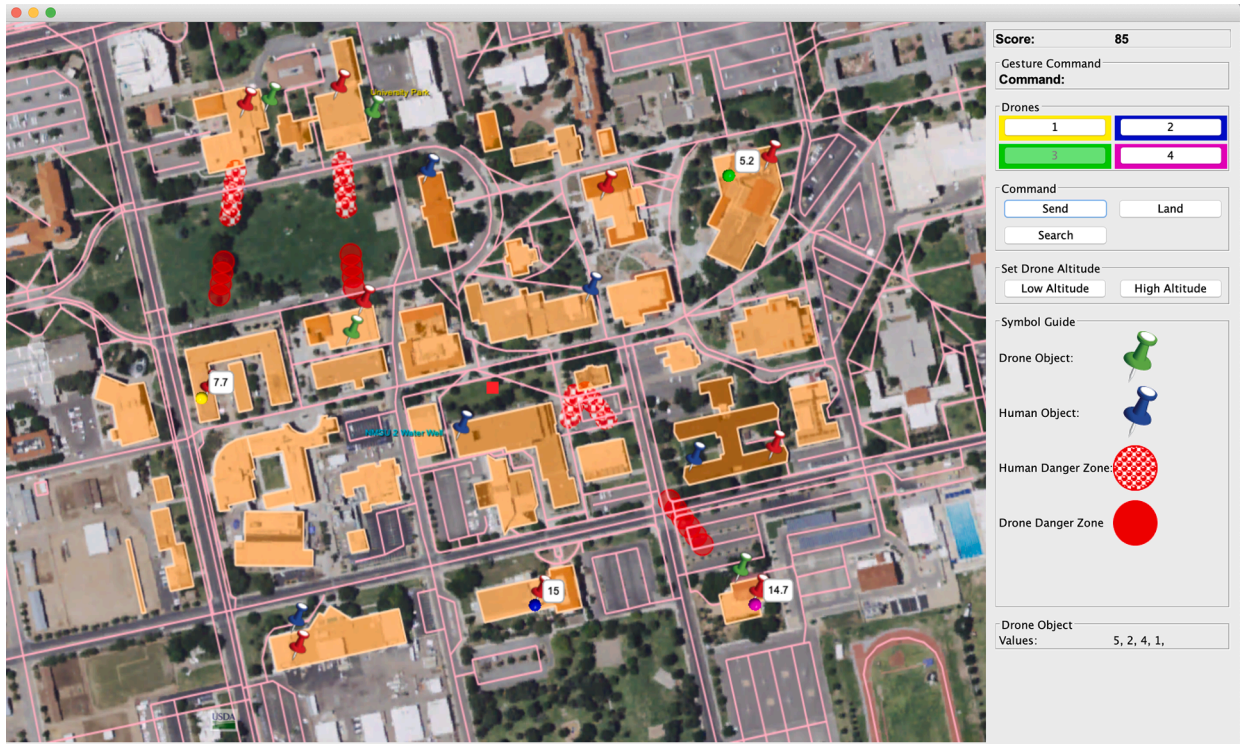


**Fig. 2.** The map UI, displayed on a touchscreen. Drones appear as a colored circle, player position is the red cube.

**Table 1**

Commands that participants can issue to drones.

| Command | Description |
|---|---|
| Send | Send drone to a new location. |
| Land | Land the drone. |
| Search | Search an area between two waypoints. |
| Low Altitude | Change drone altitude to 15m. |
| High Altitude | Change drone altitude to 25m. |

**Table 2**

Actions used by ROS for executing drone plans.

| Action | Description |
|---|---|
| Takeoff | Lift the drone off from the ground. |
| Ascend | Move the drone to a higher altitude. |
| Descend | Move the drone to a lower altitude. |
| Land | Puts the drone to the ground. (When the battery is 10%, the system automatically lands the drone.) |
| Move | Move drone from location X to location Y. |
| Scan | Drone scans the area for clues. |

**Table 3**

Virtual Drone Search Game Scoring Rubric. Game events increase or reduce player score. The highest possible score in this study is 150 points: the player must never enter a danger zone, never visit an incorrect structure, and locate and collect 10 possible goal objects.

| Action | Points |
|---|---|
| find goal object with drone | 5 |
| collect goal object with drone OR avatar | 10 |
| walk through a human danger zone | − 5 |
| visit a structure without goal object | − 2 |

attend to the wearable computer while moving in the gameworld.

Drones have limited battery power, meaning that it is not possible to do *everything* via drone. If a drone's battery runs out, the player needs to land that drone so it recharges. This also drives a need to optimally spread out drones in the environment.

Parts of terrain are dangerous either to the player or to the drones. Drones can detect these dangerous areas remotely. If a player spends time in a dangerous area, points are deducted; similarly, if a drone spends time in a dangerous area, the drone loses battery faster. In addition, if the player visits a building that does not have a hidden player object, they lose points.

Drones can be set to search an area and can locate goals within structures. If a drone goal is found, the system will display the goal location and collect it. However, if a player goal is found, the system will display the location of that object on the map; then, the player needs to move and collect the object.

The player can make choices about drone altitude, which offers trade-offs.

At low altitude, drones reveal more information to the player (i.e., danger zones, goal locations), collect drone goals, and are affected by danger zones; at high altitude, drones avoid the effects of danger zones, but cannot find and collect goal objects.

The player is free to move in the physical environment, to the limits of a specified play area. Because the game is timed, the player needs to focus on moving to the right structures with the help of intelligence provided by drones. At the same time, the player needs to identify and avoid danger zones while navigating the physical environment to structures that contain hidden objects. Ideally, the player should only search buildings that contain goal objects.

In this game, the drones had some limitations, such as battery power, but they were always accurate. We chose to make drones complete tasks accurately in this first instantiation of the game to encourage participants to use them and simplify the game during its first test. We

acknowledge that trust is important in human-machine teaming and should be further studied. However, in this paper, our primary aims are to test input modality, our system, and social perceptions. Therefore, drone accuracy and participant trust of drones is beyond the scope of this paper.

### 6.3. Wearable Interface

We developed a composite wearable UI for the game. The process of designing the wearable UI proceeded from a comprehensive analysis of existing devices to develop a framework Khalaf et al. (2020b, 2019), narrowing the design space by using the framework to identify candidate devices, then running a pilot study to verify the design. The focus for the present project is on input modalities, so we developed a set of consistent feedback modalities; testing feedback is future work.

#### 6.3.1. Grounded Theory Wearable Input Framework

To design these wearable interfaces, we undertook a grounded theory Glaser (1978, 1998); Glaser and Strauss (1967) analysis of 84 individual wearable input devices (e.g., NailO Kao et al. (2015), Tap TAP (2016)) and 197 data sources (e.g., technical specifications, research papers, instructional videos)[8]. The grounded theory analysis identified a framework to support design. We designed this system with an early version of that framework, which has with two main axes, provoking designers to consider the following questions when building composite wearable computers:

*What type of interactivity is needed / desired?* The designer should identify the TYPE OF INTERACTIVITY that is needed to accomplish the goals of the wearable system. Task demands and factors from the environment can determine what type of interactivity is most appropriate for the system. The question of *what* is used to identify the range of input that needs to be accepted by the system.

*Where might the system be worn on the body?* The designer should decide what part or parts of the body can be occupied by one or more wearable devices. The criteria for selecting the on-body location for a wearable device can be varied based on the needs of accessibility, functionality, and mobility Gemperle et al. (1998). The designer has to identify these needs to be able to determine the on-body location of the wearable device(s) and determine if devices will conflict (e.g., finger gestures may not work if the user needs to clutch a controller).

#### 6.3.2. Designing from the Framework

We used the framework to guide building a composite wearable that meets the requirements of our game. First, we need to select *what type of interactivity* the wearable should provide. The player needs to interact with the wearable while moving, based on the framework, we might use *gestures* and/or *voice*. The player needs to add waypoints on the map, so we may use *touch*, a *thumb-stick*, and/or a *trackball* to fulfill this requirement. We also can provide a *clicking* option to the player to provide their input to the system. For the mixed reality, the player needs to know their location on the map; therefore, we need to use a location tracker (GPS).

Second, we need to decide *where the system should be worn*. We want to test the effectiveness of using different wearable input devices on different parts of the body. For example, comparing between two hand gesture devices that can be worn on different parts of the body (e.g., hand versus arm). Also, We need to ensure that wearable devices do not conflict with each other. For example, players cannot wear a Myo Armband and hold Twiddler 3 in the same hand. We need to ensure that the size and weight of the wearable computer need to be appropriate for the task and will not affect player movement.

We narrowed our list to devices that provide gesture, touch, point,

---

[8] A separate manuscript in review addresses details of the analysis and resulting framework.

and click interaction and devices that would not inhibit movement in a mixed reality setting. Several devices can be used to provide gesture interaction such as Leap Motion, Intel RealSense, and Myo Armband. Leap Motion and Intel RealSense are vision-based gesture-recognition devices that can be mounted on different body parts (e.g., chest, thigh). We choose Leap Motion over other vision-based gesture-recognition sensors because it has a high level of accuracy Khalaf et al. (2020a); Weichert et al. (2013). To compare with the Leap Motion, we chose the Myo Armband, which uses a different technology to recognize gestures (inertial measurement and electromyography). As a way of interacting with a different set of gestures, we choose Tap because it combines two types of interactivity: gesture and touch; also, it fits over user fingers and would not affect player movement. To supplement, we added the Twiddler 3, a handheld device, to provide pointing and clicking capabilities. While other wearable devices also could fit our criteria, we chose these designs to minimize the number devices while meeting the criteria.

For feedback, all configurations use a wrist-worn touchscreen, which can be used as an input device, and an HMD.

### 6.3.3. Pilot Study

We ran a pilot to select which devices to use, building on data collected from our wearable framework Khalaf et al. (2020b) and our prior hand-gesture interface study Khalaf et al. (2020a). The pilot study is similar to the user study described in the next section; here we cover it briefly to support understanding the design process. All conditions (Fig. 3, left) in the pilot study used a wrist-worn touch screen for map interactions (i.e., display map, set waypoints, select drones, command drones (e.g., send, search)) and HMD for drone status (these devices are described in more detail in the final study design). The following device configurations were used for inputs in the pilot study; letters match Fig. 3, left:

(A.) Leap Motion (command drones), Twiddler (set waypoints, select drones, command drones);
(B.) Myo Armband (command drones), Twiddler (as in A.); and
(C.) Tap (set waypoints, select drones, command drones).

In the pilot study ($N = 5$; 4 female, 1 male[9]), we examined game play performance with each device configuration. At the beginning of each session, we explained the game objective and rules to each participant. Before each condition, we explained the wearable UI and how to use it; each participant got the opportunity to use the wearable UI in a tutorial.

Each participant needed over two hours to complete the study, prompting us to move to two UIs to reduce fatigue effects. The only difference between A. and B. was the gesture device; we chose to use the Leap Motion and eliminate Myo Armband because the level of accuracy of Leap Motion is higher Khalaf et al. (2020a); Weichert et al. (2013) and the Myo Armband needed a custom profile unique to each participant, which was time-consuming.

After the pilot study we modified condition C. Tap provides mouse and keyboard functionalities to the user. However, to use the mouse functionality, the player would need to place their thumb down on a flat surface and move their hand, which would be difficult in the mixed reality. Thus, we included a handheld mouse in this configuration (Fig. 3, D.).

### 6.3.4. Final Design for Study

Using data from the framework and from observations in the pilot study, we arrived at the final design of the system, which featured the two input modality combinations that serve as conditions in the study: Gesture condition (the same as pilot study A.: Fig. 3, A., Y.)) and Tap condition (Fig. 3, D., Z.). We now describe the devices that are used in both conditions and provide further detail on the two conditions.

Both conditions used the same setup, varying input modalities:

- *Touchscreen* A wrist-worn 5.8-inch touchscreen that displays the map, positions of the player and drones, and game state (e.g., danger zones, drone states; Fig. 2). The player can use the touchscreen to set drone waypoints, select a drone, and provide drone commands (e.g., change altitude, send drone).
- *HMD.* The HMD provides drone state (e.g., current action, altitude, battery level (Fig. 4).

The Gesture condition uses the following components:

- *Leap Motion:* For gesture interaction, we used the Leap Motion controller[10], a hand-gesture-recognition device that can be worn on the thigh Liu et al. (2015). We developed a simple set of gestures based on what we learned from our prior hand-gesture study Khalaf et al. (2020a). The player can use five hand gestures to provide commands (Fig. 5).
- *Twiddler:* To provide player input to the system, we used Twiddler 3[11] a handheld mouse and chording keyboard. The player can use it to set waypoints, select a drone, and issue drone commands (e.g., change altitude, send drone). The Twiddler is also used to activate gesture input by pressing a single button, preventing accidental triggering.

The Tap condition uses the following components:

- *Tap:* Tap is a hand-worn device that fits over user fingers providing mouse and keyboard inputs, able to detect hand motion and tapping on a surface. The player uses Tap to select the drones and to provide commands. To increase Tap accuracy, we customized it to recognize only 10 inputs (Table 4); a confirm command avoids spurious input.
- *Handheld Mouse:* To provide player input to the system, we used a handheld mouse that the player can use to issue drone commands.

**Gesture Mappings:**

Although prior studies have focused on optimal gesture mapping (e. g., Firestone et al. (2019)), there are not guidelines for ideal gestures for every input modality. For this study, we identified simple hand gestures that are easy for the user to perform and are recognized accurately by the device from a much earlier pilot study Khalaf et al. (2020a). We provided analogous input commands between the two wearable interfaces, as described below.

*Leap Motion:* We provided a natural mapping between the performed gesture and the command. For example, one finger, two fingers, three fingers gestures mapped to different levels of drone altitude (land, low altitude, high altitude).

*Tap:* These commands were analogous to those of Leap Motion. For example, for selecting the drone, the participant uses the same fingers in the two modalities to make the selection of the drones. For the search command, the user will use all fingers with Tap, similar to the open-hand gesture on the Leap Motion.

## 7. Method

Participants played the game two times: the two game configurations (e.g., goal positions and danger zones) were static, but different, and played in the same order, while the input modality (i.e., Gesture condition, Tap condition) was counterbalanced. We manipulated input modality within participants. The study duration was less than two hours.

---

[9] Participants had multiple options to disclose gender.

[10] https://www.LeapMotion.com
[11] https://twiddler.tekgear.com

**Fig. 3.** Left (A.–D.) are wearable UIs that we used in pilot study, a subset of which are used in the final study. A. HMD, wrist-worn touchscreen, Leap Motion, Twiddler; B. HMD, wrist-worn touchscreen, Myo Armband, and Twiddler; and C. HMD, wrist-worn touchscreen, and Tap. The modified configuration of C., used in the study: D. HMD, wrist-worn touchscreen, Tap, and handheld mouse. Right (Y., Z.) are wearable UIs in final study. Y. HMD, wrist-worn touchscreen, Leap Motion, Twiddler and Z. HMD, wrist-worn touchscreen, Tap, and a handheld mouse. All studies use a laptop computer for the gameworld UI.
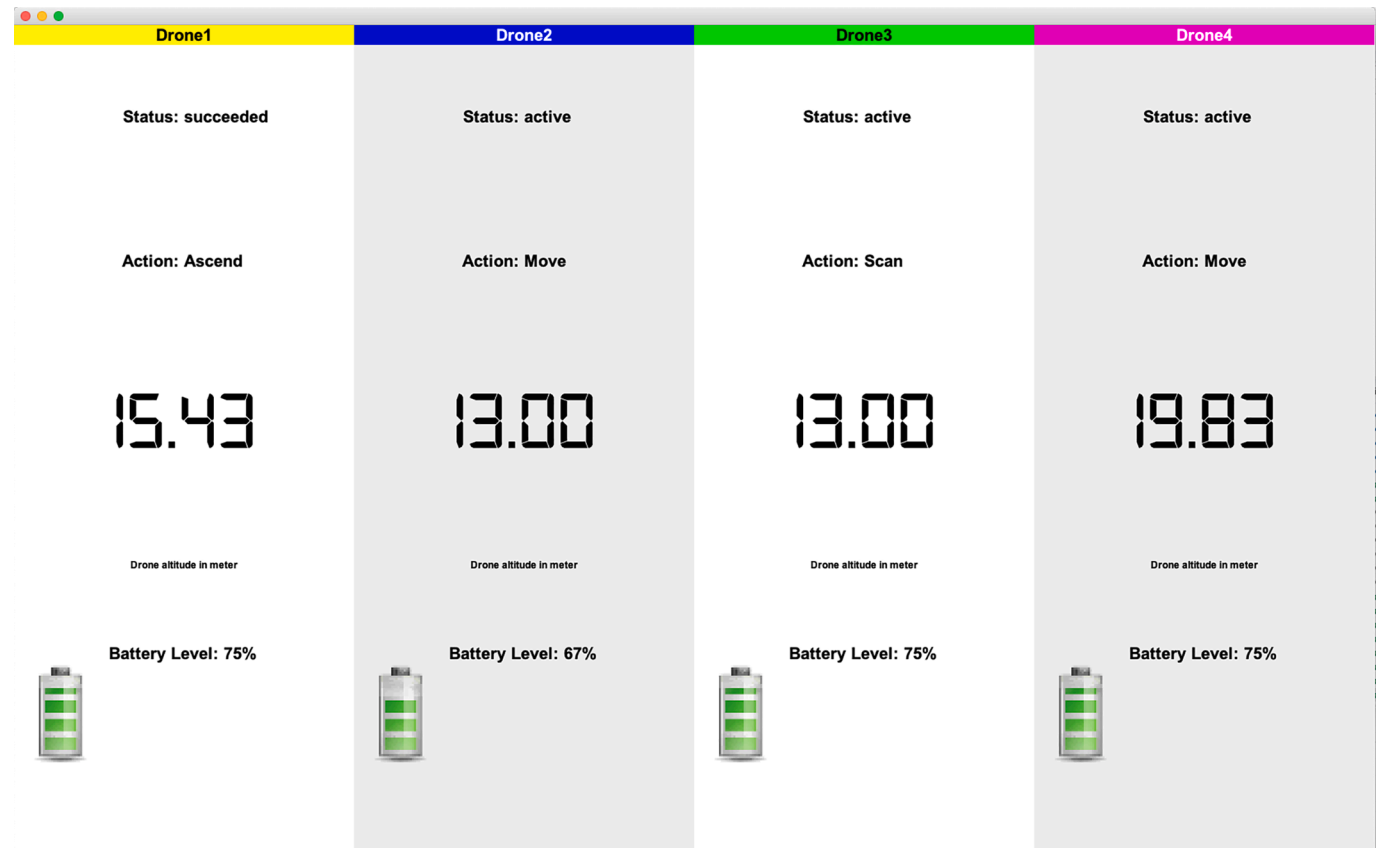


**Fig. 4.** Drone status displays, which appear on a monocular, see-through, HMD.



**Fig. 5.** Leap Motion hand gestures to direct drones. **(A)** send drone; **(B)** land drone; **(C)** change drone altitude (low altitude); **(D)** change drone altitude (high altitude); **(E)** search area.

Participants ($N = 17$; 5 female, 12 male[12]) were undergraduate students from New Mexico State University, drawn from the psychology participant pool and volunteers from computer science classes. We chose this recruitment method rather than recruiting from SAR responders in this study because SAR responders are sparse in our area and typically busy. We seek to optimize the system design in this study before including SAR responders in future studies. In this study, gender was not

---

[12] Participants had multiple options to disclose gender.

**Table 4**

Codes to direct drones: users tap the indicated finger on a surface while wearing the Tap device. Circles represent fingers; filled circles indicate which fingers trigger the command.

| Tap Code | Command | Tap Code | Command |
|---|---|---|---|
| ●○○○○ | send drone | ●●○○○ | land drone |
| ○●○○○ | select drone 1 | ○●●○○ | fly low altitude |
| ○○●○○ | select drone 2 | ○○○●● | fly high altitude |
| ○○○●○ | select drone 3 | ○●●●○ | confirm command |
| ○○○○● | select drone 4 | ●●●●● | search area |

balanced between male and female because by recruiting from within the University, we could not exclude participants based on gender. However, with our within-subjects design, we have equivalent gender proportions in the different conditions. Participants were compensated with course credit or extra credit for participation. Power analysis indicated that with a medium effect size (which we expected due to the distinct different input modalities), power would have been high with a sample of 17 participants.

### 7.1. Procedure

The study was approved by the New Mexico State University Institutional Review Board. Participants entered the lab and signed the informed consent. Then the experimenters explained the game (see description, earlier; 45 minutes), and participants took the pre-test questionnaire.

Participants played the first condition (15 minutes), which was randomly-decided. Post-test measures were collected, and participants were given up to 10 minutes to rest. Next, they played a new game (15 minutes) with different equipment in the condition they had not completed yet, and completed another post-test questionnaire.

Finally, participants completed the final questionnaires, were debriefed, and remunerated. The entire study took less than two hours.

### 7.2. Measures

We measured performance, perceived workload, situational awareness, attitude toward the drone team, general feedback, and demographics.

#### 7.2.1. Game Performance

We measured performance as the score that participants received in the game, calculated using Table 3. This indicates how effective players were at finding goals quickly, while avoiding interacting with buildings that were irrelevant.

#### 7.2.2. Workload

The *NASA Task Load Index (TLX)* is an established instrument to measure workload Hart (2006); Hart and Staveland (1988); we used it to assess the wearable UI. Participants self-reported subjective workload assessment according to the standard TLX questions, responding on hundred-point scales from 0 (not at all) to 100 (very much). The TLX was administered three times: at the beginning of the study (pre-TLX) and after each condition.

#### 7.2.3. Situational Awareness

Situational awareness is a core need in SAR Seppanen et al. (2013), so we test it in our design. We measured situational awareness using a method similar to situational awareness Global Assessment Technique (SAGAT) Endsley (2000). SAGAT is normally used in larger scale simulations and uses more deeply developed, mission-specific questions and involves pausing the simulation. We asked participants simpler questions and did so at the end of each condition. We avoided any increase in

strain on memory by asking participants about awareness of the situation as it was immediately before the condition ended – thus making it consistent with typical SAGAT measures.

Participants were asked the approximate location of each of the four drones at the end of their game. We chose location because it is critical for users to be aware of drones' locations so they can quickly determine how to direct them. We chose drone location in relation to the overall map because it related to participants' goal of exploring the overall map for clues. Participants saw and selected from the same map they were using divided into a $3 \times 3$ grid labeled 1–9. Answers were marked correct if participants chose the correct ninth of the map.

We also asked participants the percent battery remaining each drone had after each condition. We chose this because when a drone's battery dies, it must land to recharge. By being aware of drone battery, participants had the possibility of strategically flying and landing drones to efficiently coordinate search efforts. The experimenter recorded the actual percent battery remaining for each drone after each condition. A measure of situational awareness was created by taking the absolute value of the participants' answer minus the actual answer for each drone, then averaging the measure for each drone per condition. Overall, this created a measure of on average how close participants were to remembering the actual battery life of each drone per condition.

#### 7.2.4. Attitude toward the Drone Team

We assessed players' self-reported perceptions of their drone teammates.

*Entitativity.* We measured perceived entitativity using four questions from several scales from 1 (not at all) to 9 (very much) Castano et al. (2003a); Kurebayashi et al. (2012); Lickel et al. (2001b); Rydell et al. (2007); Rydell and McConnell (2005); Yuki (2003). Questions were averaged to measure overall entitativity (e.g., "This group of robots should be thought of as a whole"). We measured entitativity after each condition.

*Emotions.* We measured emotion toward robots from 1 (not at all) to 7 (very much; e.g., "I fear robots"; "I respect robots") Cottrell and Neuberg (2005b). We measured emotion after each condition.

*Anthropomorphism.* We measured perceptions of drone anthropomorphism with Kozak's scale Kozak et al. (2006) from 1 (not at all) to 7 (very much). The measure includes two subscales: agency (e.g., these robots "have goals") and experience (e.g., these robots "are capable of emotion"). We measured this once, at the end of the study.

#### 7.2.5. General Feedback

Participants responded to free response questions of their experience of the computer system and user interfaces (e.g., "What kinds of questions or difficulties did you have when playing?"). Participants were briefly interviewed after each condition.

#### 7.2.6. Demographics.

We documented demographics (age, gender, field of study, technology experience). We measured this once, at the end.

## 8. Results

Data were analyzed in JASP version 9.2[13] and SPSS version 25 for measures that used Cronbach's alpha or factor analysis. Data met assumptions for the tests we ran. We ran Kurtosis and Skewness tests of normality. If the original data did not meet assumptions, we describe in that section how we first normalized the data. We considered values of $p < .05$ to be statistically significant. All significant effects are reported. For each of the research questions, we report on relevant measures.

---

[13] JASP https://jasp-stats.org/

### 8.1. RQ 1: Input Modality Effectiveness

#### 8.1.1. Performance

Condition affected score ($F(1, 15) = 16.847$, $p = .001$, $\eta_p^2 = .518$), with participants performing better using Gesture than Tap (Table 6). There was a significant interaction effect of Condition and Sequence ($F(1, 15) = 3.798$, $p = .070$, $\eta_p^2 = .202$): participants performed better in the second game.

Despite our best efforts to make the game long enough that all participants could not finish, three participants (split across conditions) finished early. However, there is no statistically significant difference in time finished across conditions or order of conditions.

#### 8.1.2. Situational Awareness

Participants performed very poorly on the situational awareness measure of drone location, getting approximately 15% of the four questions correct. If they had been guessing by chance, they would have had 11% correct on each question. An ANOVA indicated no statistically significant differences across conditions (Table 5). However, participants performed fairly well on situational awareness measure of drone battery life, coming within approximately 10% from the correct answer. A repeated measures ANOVA revealed that this did not differ significantly across conditions.

#### 8.1.3. TLX

We performed 3 (Condition: Before, Gesture, Tap) ×2 (Sequence: Gesture-first, Tap-first) ANOVAs on the TLX data. We used posthoc tests with Bonferoni corrections to differentiate significant differences. Below, we break down findings by TLX category (see Table 6 and Fig. 7).

*Mental Demand.* Condition affected perceived mental demand ($F(2, 30) = 5.871$, $p = .007$, $\eta_p^2 = .270$) such that participants rated the Tap condition as significantly more mentally demanding than the Before condition ($p = .009$).

*Physical Demand.* Condition affected perceived physical demand ($F(2, 30) = 4.178$, $p = .025$, $\eta_p^2 = .202$), but posthoc tests indicated no significant differences between conditions. Sequence affected physical demand ($F(1, 15) = 4.701$, $p = .047$, $\eta_p^2 = .239$) such that participants reported greater physical work load in Gesture-first than Tap-first.

*Temporal Demand.* Condition affected perceived temporal demand ($F(2, 30) = 5.495$, $p = .009$, $\eta_p^2 = .249$) such that participants indicated being more rushed in the Tap than the Before condition ($p = .033$).

*Performance.* Condition affected perceived performance ($F(2, 30) = 9.156$, $p < .001$, $\eta_p^2 = .287$) such that participants rated the Tap condition as characterized by significantly higher perceived failure than the Gesture condition ($p = .006$). An interaction effect between Condition and Sequence ($F(2, 30) = 7.701$, $p = .002$, $\eta_p^2 = .242$) indicated that participants rated sessions as having more failure in the condition they performed second.

*Effort.* Condition affected perceived effort ($F(2, 30) = 25.511$, $p = .001$, $\eta_p^2 = .612$) such that participants indicated more effort in the Tap than the Gesture ($p = .001$) and the Before ($p = .001$) conditions.

*Frustration.* Conditions did not affect frustration.

### 8.2. RQ2: Improving the System

#### 8.2.1. General Feedback

For general feedback, we performed a lightweight thematic analysis Braun and Clarke (2006) on the responses. One researcher read the responses and developed themes for the most common responses: drone battery/charging, where things are, moving the avatar, controls (including finger motions, drone control, remembering commands), the game itself, and other. Two researchers used the initial themes to independently code the data. There was substantial agreement between coders ($\kappa = .756$, $p < .001$) Landis and Koch (1977).

For the themes that coders disagreed on, they discussed how responses should be properly categorized. Most items that were disagreed upon did not fit into any category because the context of the response was unclear, (e.g., "There are many directions,") or it applied to both controls and the game (e.g., "multitasking," "finding clues"). To handle this, we created a new theme corresponding to multitasking and finding clues, which related to both the device and the game.

Answers to each question were coded the same way, so we report overall what problems participants had (Table 7). Across conditions, participants reported the most confusion about drone control. Participants in the Gesture condition reported more confusion about multitasking/finding clues, and participants in the Tap condition reported more with the game. Participants in the Tap-first condition reported more confusion about drone battery/charging and the game than Gesture-first. According to [P22], "The most confusing part of the game was getting the hang of all the tools I had to use; the hardest was the hand control that directed where the drones went."

Two specific aspects of game control that most confused and challenged participants were the drone battery/charging and moving the avatar. One participant said, "I forgot to keep track of battery and done location" [P20]. As first-time players, participants had trouble remembering most of the hand gesture commands and how to control the drones. A comment from one of our participants illustrates some of these issues: "One of the challenging parts of the game is to remember the commands" [P7].

#### 8.2.2. Observations

Although participants did not report this in the general feedback, several participants across conditions stated to the experimenters that it was hard for them to use the phone-sized touchscreen because the screen was too small to see much of the map. They requested a larger screen if possible. Also, some participants suggested putting the map screen on the HMD so that they can see the pins on the map closer: "Perhaps you could place the battery screen on the phone and the map screen on the HMD. I feel that the game would be easier to play that way" [P8].

### 8.3. RQ3: Social Experience of (Virtual) Drones

#### 8.3.1. Entitativity

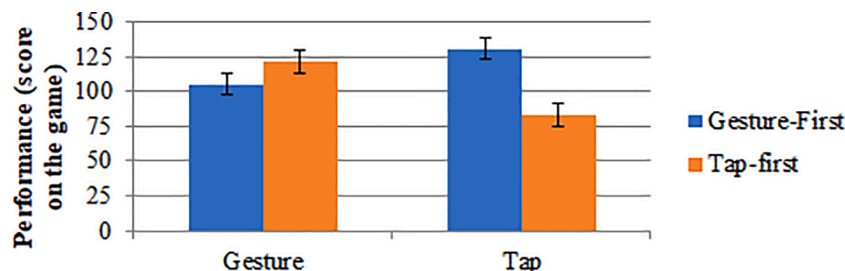Entitativity scales had high Cronbach's $\alpha$ (Robot Group = .841,



**Fig. 6.** Participant performance as measured by score in the game. Error bars denote standard error.
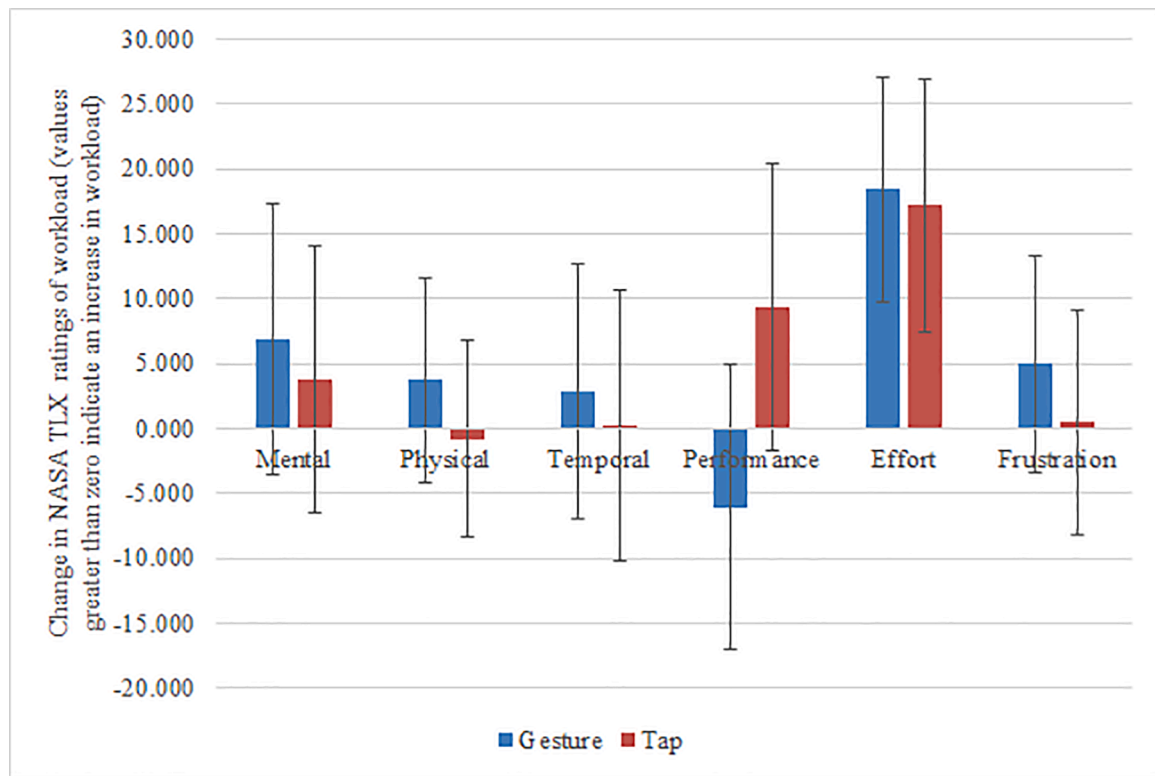
**Fig. 7.** Participant ratings of workload on the NASA TLX. Because there were no significant effects of order, we collapse across order. This figure shows the difference between the before measure and each condition. Error bars denote standard error.

**Table 5**
Performance and situational awareness (SA) across conditions. Presented as $M(SD)$.

| | Gesture | | Tap | |
|---|---|---|---|---|
| | Gesture-first | Tap-first | Gesture-first | Tap-first |
| SA (Location (of 4)) | 1.00 (.82) | 0.71 (.76) | 1.17 (.75) | 0.50 (.55) |
| SA (Battery (of 100%)) | 12.61 (7.95) | 9.56 (8.74) | 11.82 (12.40) | 8.56 (8.41) |

Human & Robot Group = .810).

On average, people viewed themselves as slightly more like a group than neutral A $2 \times 2$ repeated measures ANOVA showed no significant differences across conditions (Table 9).

*8.3.2. Emotions*

A factor analysis revealed three factors: Negative (anger, fear, disgust, resentment, anxiety, uneasiness, sadness; Cronbach's $\alpha = .930$), Sympathy (pity, guilt, sympathy; $\alpha = .897$), and Positive (respect, happiness, gratefulness, excitement; $\alpha = .818$). Participants reported very low levels of negative and sympathetic emotions, and to

moderately experience positive emotions, toward the robots (Table 8). Negative and sympathetic emotions were skewed right (skewness = 3.161 and 2.788, respectively). For running statistical tests, used the natural log of the data, which reduced skewness to the acceptable level of between -2 and 2 (skewness = 1.876 and 1.489, respectively. A series of $2 \times 2$ repeated measures ANOVAs revealed no statistically significant effect of condition or sequence on emotions.

*8.3.3. Anthropomorphism*

On anthropomorphism measures (taken once per participant), participants rated robots as very low on experience (Gesture-first $M = 2.54$, $SD = 1.00$, Tap-first $M = 3.24$, $SD = 1.60$), and slightly below moderate on agency (Gesture-first $M = 1.17$, $SD = 0.36$, Tap-first $M = 1.63$, $SD = 0.73$). T-tests indicated no significant differences across sequence.

**9. Discussion**

In this experiment, participants played the *Virtual Drone Search Game*, controlling an avatar from a first-person perspective and directing four drones with an overhead perspective. We compared performance and workload between two methods of controlling the drone: Gesture and Tap. Overall, participants performed better and experienced less workload using Gesture than Tap (RQ1). Qualitative results indicated

**Table 6**
Exact value of participant ratings of workload on the NASA TLX (Fig. 7). Presented as $M(SD)$.

| | Before | | Gesture | | Tap | |
|---|---|---|---|---|---|---|
| | Gesture-first | Tap-first | Gesture-first | Tap-first | Gesture-first | Tap-first |
| Mental | 29.44 (28.62) | 39.75 (33.06) | 65.22 (30.82) | 53.50 (29.76) | 59.00 (30.88) | 63.13 (24.63) |
| Physical | 21.22 (32.03) | 13.50 (17.55) | 48.44 (27.46) | 21.00 (12.82) | 39.33 (27.96) | 15.50 (12.66) |
| Temporal | 10.11 (12.35) | 21.75 (30.06) | 36.78 (31.56) | 27.50 (30.07) | 31.56 (27.31) | 41.50 (30.13) |
| Performance | 89.78 (15.48) | 81.25 (37.20) | 47.00 (34.62) | 69.13 (32.30) | 77.78 (31.22) | 45.13 (22.15) |
| Effort | 19.11 (19.38) | 9.25 (10.05) | 65.67 (26.28) | 46.13 (30.74) | 63.11 (32.90) | 66.75 (23.08) |
| Frustration | 16.22 (27.57) | 12.75 (23.71) | 20.00 (20.11) | 22.75 (27.32) | 11.00 (19.17) | 37.00 (38.07) |

**Table 7**

General feedback themes addressing what was confusing, divided across conditions; numbers represent counts of each theme and example for each theme.

| | Example | Gesture | | Tap | | Total total |
|---|---|---|---|---|---|---|
| | | Gest. first | Tap first | Gest. first | Tap first | |
| **Drone battery/ charging** | "I forgot to keep track of battery and done location." | 1 | 6 | 1 | 4 | 12 |
| **Where things are** | "Knowing which buildings had the pins…" | 1 | 1 | 1 | 1 | 4 |
| **Moving avatar** | "Directing the person around the buildings…" | 3 | 4 | 5 | 3 | 15 |
| **Drone control** | "Learning the finger motions used to control the drones." | 19 | 11 | 13 | 16 | 59 |
| **Game** | "Finding every clue on the map." | 1 | 5 | 5 | 4 | 15 |
| **Multitasking/ finding clues** | "Keeping track of all the tasks on the screen." | 5 | 5 | 2 | 4 | 16 |
| **Uncategorizable** | "Remembering the overall objective of the game." | 2 | 2 | 5 | 2 | 11 |

**Table 8**

Emotions rated toward the drones on a 1–7 Likert scale, reported as $M(SD)$.

| | Gesture | | Tap | |
|---|---|---|---|---|
| | Gesture-1st | Tap-1st | Gesture-1st | Tap-1st |
| **Negative** | 1.73 (0.89) | 1.18 (1.22) | 1.61 (0.32) | 1.33 (0.50) |
| **Sympathy** | 1.46 (0.67) | 1.67 (0.79) | 1.46 (0.73) | 1.89 (1.79) |
| **Positive** | 3.16 (1.64) | 3.50 (1.84) | 2.97 (1.45) | 3.28 (1.34) |

**Table 9**

Entitativity rated toward the robot group and toward the human group, reported as $M(SD)$.

| | Gesture | | Tap | |
|---|---|---|---|---|
| | Gesture-1st | Tap-1st | Gesture-1st | Tap-1st |
| **Entitativity toward robot group** | 5.00 (0.94) | 6.28 (2.07) | 4.53 (1.92) | 6.28 (1.54) |
| **Entitativity toward human group** | 4.81 (1.24) | 5.83 (2.32) | 4.91 (1.92) | 6.28 (1.54) |

that the most challenging part of both studies was the multiple controls (RQ2). Finally, participants felt like a group with the drones. They showed low levels of negative and sympathetic emotions and moderate levels of positive emotions toward the drones. They did not identify the drones as particularly anthropomorphic (RQ3). These results are discussed in more detail below.

### 9.1. RQ1: Which input modality will work more effectively?

In response to RQ1, the Gesture condition was better across measures than the Tap condition. Participants had higher performance using Gesture, and rated Gesture as more positive on the TLX (less mental demand, less hurry, more success, less workload) than Tap. There was no difference across conditions for situational awareness. These effects were robust, occurring despite the (counterbalanced) order effects of participants performing better the second time they completed the task. One reason for the greater difficulty of the Tap than Gesture condition may be the calibration of the tools. Both devices recognize gesture, but

the Gesture system we used had been tested in the past, resulting in high levels of accuracy compared to other gesture devices. Also, in the Gesture condition, players used the devices to perform only five commands and used the Twiddler to select drones. In the Tap condition, the player used the devices to perform the five commands as well as to select the four drones. Adding many ways to use this technology may have increased confusion and decreased performance. Also, the user might forget the commands for the Tap. In the future, it may be preferred to use Gesture over Tap for controlling multiple drones. Researchers and designers should also minimize the number of commands drone users need to use and memorize, especially for first-time users.

### 9.2. RQ2: How can we improve our system?

Future studies should include an early training session with participants on the technology. Most participants were confused or had difficulties related to the control of drones, including gestures and using the equipment because they had no background in using the devices. Future studies could also examine how long it takes participants to reach a sufficient level of expertise for the devices, to provide an estimate of how long SAR responders should be trained with this type of wearable technology.

Related, participant strategies likely change over the course of the game as they understand the game better. About one third of comments related to confusion about the game, finding clues, or its difficulty, but participants showed improvement from their first to second game, both in quantitative performance and in qualitative comments. Because SAR respondents are already familiar with the process of SAR, performance on this game may relate most to SAR responder performance after participants have developed an understanding of the game. Future studies might include a 10-minute training session or divide the game into segments to assess performance as participants become familiar with the game.

It might help to rework how the drone battery/charging functions or pilot test a more helpful explanation. Confusion about the avatar was likely because this prototype was played on a computer screen; we expect decreased confusion in a mixed-reality game, because players will simply be in the physical world. Further, participants likely paid little attention to the first-person gameworld UI, and more to the drones, because they could move the avatar without worrying about real life complications like tripping or running into buildings. Future studies with this gameworld UI should add reasons for participants to pay closer attention to the avatar screen, such as game points detracted if participants "trip," a fall that could be avoided by paying close attention for game obstacles.

Participants indicated difficulty using the small touch screen. Future studies might use a wearable tablet, giving participants more real estate for directing drones. Using the HMD to display the game map instead of the drone battery could provide participants with more explicit and closer views.

### 9.3. RQ3: What is a baseline for social experience of drones?

Related to RQ3, overall, participants felt like a group with the simulated drones to a moderate extent. They did not feel that the drones were a group without them any more than that they were a group with the drones. This is interesting because prior work in social psychology indicates that when people feel like a group with others, they are more likely to work with the group again Campbell (1958); Lickel et al. (2001a). The results suggest that participants would work with this group of simulated drones again.

Participants rated low levels of negative and sympathetic emotions toward the simulated drones and a moderate level of positive emotions toward them, indicating that they neither particularly liked nor disliked the drones. In terms of anthropomorphism, participants rated the drones as agentic, but with little ability to experience positive or negative

stimuli. These results will inform future studies; as interaction schemes with the drones change, participants may come to view the drones as increasingly more capable of acting on the world and perhaps even capable of experiencing.

### 9.4. Limitations and Future Directions

Like any study, the study has its own set of limitations. Participants in this study were psychology and computer science students with little experience with these technologies. This means that these results might extend to first-time users, whereas if this technology is used consistently, long-term benefits of using one technology over another may differ. Participants may also have had different survey responses on sociality scales than non-students or SAR responders.

In future work, we recommend increasing the number of participants and including participants from SAR and others who have real experience in high-stress situations. We also suggest a visit to a SAR training facility, involving touring the training ground and conducting in-depth face-to-face interviews with SAR responders, to understand planning, practice, and training methods.

Further, we performed the study in the lab on computers. The situation differs from the real world (e.g., participants were not walking and paying attention to the physical world while directing drones). In future work, we expect to design outdoor mixed reality environments that combine virtual reality information with physical reality experience.

In this study, participants had low uncertainty related to drones providing accurate information. Future studies should examine how greater uncertainty of drone performance would affect participant use of the wearable devices and system. Differing levels of drone autonomy would also affect our measures of entitativity, emotion, and anthropomorphism, and should be examined in future studies. In this study, the drones had limited to no autonomy, but drones of higher levels of autonomy would be perceived as more anthropomorphic Haslam et al. (2008); Kahn Jr. et al. (2012), and people would likely have stronger emotions toward them Kahn Jr. et al. (2012); Lee and Lau (2011); Waytz et al. (2010).

## 10. Conclusion

In this study, participants controlled multiple drones using Gesture and Tap input modalities. We found increased performance and decreased workload for the Gesture compared to the Tap condition. We recommend that future studies test Tap technology in a more real-world simulation.

In future work, we will deploy this game in a more real-world setting, with players moving through the physical world with teams of virtual and potentially real drones. Thus, many elements that are not specific to the virtual gameworld are built such that they can readily apply to real-world contexts. PDDL, ROS, and Gazebo are production-ready platforms – that is, the code could be installed directly on physical drones and run in the physical world, using real sensors instead of those simulated by Gazebo. The wearable UI, in the form of the composite wearable computer and its software, takes data directly from this platform and is agnostic to the game and centered around expected SAR practice and drone control. With a set of hardware drones and a space to legally fly them, the virtual gameworld components could be turned off and the whole system run as-is. The design begins to address SAR methods, with players faced with decisions about performing activities themselves or having a drone perform them. As this system improves, it may also be helpful for other safety-critical applications, like wild fire fighting and teaming in factories.

## CRediT authorship contribution statement

**Marlena R. Fraune:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Ahmed S. Khalaf:** Methodology, Software, Validation, Investigation, Data curation, Writing - original draft. **Mahlet Zemedie:** Formal analysis, Investigation, Data curation, Writing - original draft. **Poom Pianpak:** Methodology, Software, Validation, Investigation, Data curation, Writing - review & editing. **Zahra NaminiMianji:** Methodology, Software, Writing - original draft. **Sultan A. Alharthi:** Methodology, Software, Writing - original draft. **Igor Dolgov:** . **Bill Hamilton:** Conceptualization, Investigation, Writing - review & editing. **Son Tran:** Conceptualization, Funding acquisition, Writing - review & editing. **Phoebe O. Toups Dugas:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they do not have any financial or nonfinancial conflict of interests

## Acknowledgments

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ijhcs.2020.102573

## References

ABC News, 2017. First responders use drones in harvey rescues. Video.

Adams, E., Dormans, J., 2012. Game Mechanics: Advanced Game Design, 1st. New Riders Publishing, Thousand Oaks, CA, USA.

Alharthi, S.A., LaLone, N., Khalaf, A.S., Torres, R., Nacke, L., Dolgov, I., Toups Dugas, P.O., 2018a. Practical insights into the design of future disaster response training simulations. In: Boersma, K., Tomaszeski, B. (Eds.), ISCRAM 2018 Conference Proceedings – 15th International Conference on Information Systems for Crisis Response and Management, pp. 818–830.

Alharthi, S.A., Sharma, H.N., Sunka, S., Dolgov, I., Toups Dugas, P.O., 2018b. Designing Future Disaster Response Team Wearables from a Grounding in Practice. Proceedings of Technology, Mind, and Society. ACM, New York, NY, USA.

Ashbrook, D., Lyons, K., Clawson, J., Starner, T., 2009. Methods of evaluation for wearable computing. In: Cho, G. (Ed.), Smart Clothing: Technology and Applications, Human Factors and Ergonomics Series. CRC Press, pp. 229–248.

Austin, R., 2010. Unmanned aircraft systems: UAVS design, development and deployment, 54. John Wiley & Sons, Ltd.

Barfield, W., 2015. Fundamentals of wearable computers and augmented reality. CRC Press.

Barin, A., Dolgov, I., Toups Dugas, P.O., 2017. Understanding dangerous play: A grounded theory analysis of high-performance drone racing crashes. Proceedings of the Annual Symposium on Computer-Human Interaction in Play. ACM, New York, NY, USA, pp. 485–496. https://doi.org/10.1145/3116595.3116611.

Bellotti, V., Edwards, K., 2001. Intelligibility and accountability: Human considerations in context-aware systems. Hum.-Comput. Interact. 16 (2), 193–212. https://doi.org/10.1207/S15327051HCI16234_05.

Bevacqua, G., Cacace, J., Finzi, A., Lippiello, V., 2015. Mixed-initiative planning and execution for multiple drones in search and rescue missions. Twenty-Fifth International Conference on Automated Planning and Scheduling.

Brambilla, M., Ferrante, E., Birattari, M., Dorigo, M., 2013. Swarm robotics: A review from the swarm engineering perspective. Swarm Intelligence 7 (1), 1–41. https://doi.org/10.1007/s11721-012-0075-2.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3 (2), 77–101.

Cacace, J., Finzi, A., Lippiello, V., 2016. Multimodal interaction with multiple co-located drones in search and rescue missions. arXiv preprint arXiv:1605.07316.

Cacace, J., Finzi, A., Lippiello, V., 2017. Robust multimodal command interpretation for human-multirobot interaction. AIRO@ AI* IA, pp. 27–33.

Campbell, D.T., 1958. Common fate, similarity, and other indices of the status of aggregates of persons as social entities. Behavioral Science 3 (1), 14–25.

Carpenter, J., 2016. Culture and human-robot interaction in militarized spaces: A war story. Routledge.

Castano, E., Sacchi, S., Gries, P.H., 2003a. The perception of the other in international relations: Evidence for the polarizing effect of entitativity. Political Psychology 24 (3), 449–468. https://doi.org/10.1111/0162-895X.00336.

Castano, E., Yzerbyt, V., Bourguignon, D., 2003b. We are one and i like it: The impact of ingroup entitativity on ingroup identification. European journal of social psychology 33 (6), 735–754.

Chang, V., Chundury, P., Chetty, M., 2017. Spiders in the sky: User perceptions of drones, privacy, and security. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 6765–6776. https://doi.org/10.1145/3025453.3025632.

Chao, H., Cao, Y., Chen, Y., 2010. Autopilots for small unmanned aerial vehicles: A survey. International Journal of Control, Automation and Systems 8 (1), 36–44.

Chen, J.Y.C., Barnes, M.J., 2012. Supervisory control of multiple robots: effects of imperfect automation and individual differences. Human Factors: The Journal of the Human Factors and Ergonomics Society 54 (2), 157–174.

Chien, S.-Y., Wang, H., Lewis, M., 2010. Human vs. algorithmic path planning for search and rescue by robot teams. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 54 (4), 379–383. https://doi.org/10.1177/154193121005400423.

Chien, S.-Y., Wang, H., Lewis, M., Mehrotra, S., Sycara, K., 2011. Effects of alarms on control of robot teams. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55. Sage Publications, pp. 434–438.

Cottrell, C.A., Neuberg, S.L., 2005a. Different emotional reactions to different groups: a sociofunctional threat-based approach to" prejudice". Journal of Personality and Social Psychology 88 (5), 770.

Cottrell, C.A., Neuberg, S.L., 2005b. Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice". Journal of Personality and Social Psychology 88 (5), 770–789. https://doi.org/10.1037/0022-3514.88.5.770.

Dasgupta, N., Banaji, M.R., Abelson, R.P., 1999. Group entitativity and group perception: Associations between physical features and psychological judgment. Journal of Personality and Social Psychology 77 (5), 991.

Department of Defense, 2006. *DoD Support to Civil Search and Rescue (SAR)*. https://www.hsdl.org/?view&did=459986.

Dolgov, I., Hottman, S.B., 2011. Human factors in unmanned aircraft systems. In: Barnhart, R.K., Hottman, S.B., Marshall, D.M., Shappee, E. (Eds.), Introduction to Unmanned Aircraft Systems. CRC Press, pp. 165–180. https://doi.org/10.1201/b11202-12.

Dolgov, I., Kaltenbach, E., Khalaf, A.S., Toups Dugas, P.O., 2017. Measuring human performance in the field: Concepts and applications for unmanned aircraft systems. In: Cuevas, H., Velasquez, J., Dattel, A. (Eds.), Human Factors in Practice: Concepts and Applications. Taylor & Francis.

Dorr, L., 2018. Fact sheet – small unmanned aircraft regulations (part 107). Fact Sheet.

Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Human Factors 37 (1), 32–64.

Endsley, M.R., 2000. Theoretical underpinnings of situation awareness: A critical review. In: Endsley, M.R., Garland, D.J. (Eds.), Situation Awareness Analysis and Measurement. Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp. 3–6.

Epley, N., Waytz, A., 2010. Mind perception. In: Fiske, S.T., Gilbert, D.T., Lindzey, G. (Eds.), Handbook of Social Psychology. John Wiley & Sons, Ltd, pp. 498–541.

Epley, N., Waytz, A., Cacioppo, J.T., 2007. On seeing human: a three-factor theory of anthropomorphism. Psychological review 114 (4), 864.

Epson, 2016. *Moverio smart glasses tested in disaster response system*. https://www.epson.co.uk/insights/casestudy/moverio-smart-glasses-tested-in-disaster-response-system.

Feese, S., Arnrich, B., Tröster, G., Burtscher, M., Meyer, B., Jonas, K., 2013. Sensing Group Proximity Dynamics of Firefighting Teams Using Smartphones. Proceedings of the 2013 International Symposium on Wearable Computers. ACM, New York, NY, USA, pp. 97–104. https://doi.org/10.1145/2493988.2494332.

Firestone, J.W., Quiñones, R., Duncan, B.A., 2019. Learning from users: an elicitation study and taxonomy for communicating small unmanned aerial system states through gestures. 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, pp. 163–171.

Fischer, J.E., Greenhalgh, C., Jiang, W., Ramchurn, S.D., Wu, F., Rodden, T., 2017. In-the-loop or on-the-loop? interactional arrangements to support team coordination with a planning agent. Concurrency and Computation: Practice and Experience e4082–n/a. https://doi.org/10.1002/cpe.4082.E4082 cpe.4082

Fischer, J.E., Jiang, W., Kerne, A., Greenhalgh, C., Ramchurn, S., Reece, S., Pantidi, N., Rodden, T., 2014. Supporting Team Coordination on the Ground: Requirements from a Mixed Reality Game. In: Rossitto, C., Ciolfi, L., Martin, D., Conein, B. (Eds.), COOP 2014 - Proceedings of the 11th International Conference on the Design of Cooperative Systems. Springer International Publishing, pp. 49–67. https://doi.org/10.1007/978-3-319-06498-7_4.

Fischer, J.E., Reeves, S., Rodden, T., Reece, S., Ramchurn, S.D., Jones, D., 2015. Building a birds eye view: Collaborative work in disaster response. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 4103–4112. https://doi.org/10.1145/2702123.2702313.

Fraune, M.R., Šabanović, S., Smith, E.R., 2017a. Teammates first: Favoring ingroup robots over outgroup humans. 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, pp. 1432–1437.

Fraune, M.R., Šabanović, S., Smith, E.R., Nishiwaki, Y., Okada, M., 2017b. Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, pp. 205–213.

Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., Martin, R., 1998. Design for wearability. Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215), pp. 116–122. https://doi.org/10.1109/ISWC.1998.729537.

Glaser, B.G., 1978. Theoretical Sensitivity: Advances in the Methodology of Grounded Theory. The Sociology Press, Mill Valley, California, USA.

Glaser, B.G., 1998. Doing Grounded Theory: Issues and Discussions. The Sociology Press, Mill Valley, California, USA.

Glaser, B.G., Strauss, A., 1967. The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing Company, Chicago, Illinois, USA.

Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. Science 315 (5812).619–619.

Hart, S.G., 2006. Nasa-task load index (nasa-tlx); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50 (9), 904–908. https://doi.org/10.1177/154193120605000909.

Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload, Advances in Psychology, 52. North-Holland, Amsterdam, Netherlands, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Haslam, N., Holland, E., Karasawa, M., 2013. Essentialism and entitativity across cultures. Culture and Group Processes 17–37.

Haslam, N., Loughnan, S., Kashima, Y., Bain, P., 2008. Attributing and denying humanness to others. European Review of Social Psychology 19 (1), 55–85.

Hayes, S., Hooten, E., Adams, J., 2010. Multi-touch interaction for tasking robots. Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, pp. 97–98. https://doi.org/10.1109/HRI.2010.5453253.

Heinssen Jr., R.K., Glass, C.R., Knight, L.A., 1987. Assessing computer anxiety: Development and validation of the computer anxiety rating scale. Computers in Human Behavior 3 (1), 49–59.

Hobbs, A., 2010. Unmanned aircraft systems. In: Salas, E., Maurino, D. (Eds.), Human Factors in Aviation, 2nd. New York: Elsevier, pp. 505–531. https://doi.org/10.1016/B978-0-12-374518-7.00016-X.

Hsu, E.B., Li, Y., Bayram, J.D., Levinson, D., Yang, S., Monahan, C., 2013. State of virtual reality based disaster preparedness and response training. PLoS Currents 5.

Hutson, M., 2017. *Hurricanes Show Why Drones Are the Future of Disaster Relief*. https://www.nbcnews.com/mach/science/hurricanes-show-why-drones-are-future-di saster-relief-ncna799961.

Insko, C.A., Wildschut, T., Cohen, T.R., 2013. Interindividual–intergroup discontinuity in the prisoner's dilemma game: How common fate, proximity, and similarity affect intergroup competition. Organizational Behavior and Human Decision Processes 120 (2), 168–180.

Jones, B., Dillman, K., Tang, R., Tang, A., Sharlin, E., Oehlberg, L., Neustaedter, C., Bateman, S., 2016. Elevating communication, collaboration, and shared experiences in mobile video through drones. Proceedings of the 2016 ACM Conference on Designing Interactive Systems. ACM, New York, NY, USA, pp. 1123–1135. https://doi.org/10.1145/2901790.2901847.

Jørgensen, K., 2013. Gameworld Interfaces. MIT Press, Cambridge, MA, USA.

Juul, J., 2005. Half Real: Video Games between Real Rules and Fictional Worlds. MIT Press, Cambridge, MA, USA.

Kahn Jr., P.H., Kanda, T., Ishiguro, H., Freier, N.G., Severson, R.L., Gill, B.T., Ruckert, J.H., Shen, S., 2012. "robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. Developmental Psychology 48 (2), 303–314.

Kao, H.-L.C., Dementyev, A., Paradiso, J.A., Schmandt, C., 2015. Nailo: Fingernails as an input surface. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 3015–3018. https://doi.org/10.1145/2702123.2702572.

Karaca, Y., Cicek, M., Tatli, O., Sahin, A., Pasli, S., Beser, M.F., Turedi, S., 2018. The potential use of unmanned aircraft systems (drones) in mountain search and rescue operations. The American Journal of Emergency Medicine 36 (4), 583–588.

Khalaf, A.S., Alharthi, S.A., Alshehri, A., Dolgov, I., Toups Dugas, P.O., 2020a. A comparative study of hand-gesture recognition devices for games. Lecture Notes in Computer Science 12182, 57–76. https://doi.org/10.1007/978-3-030-49062-1_4.

Khalaf, A.S., Alharthi, S.A., Hamilton, B., Dolgov, I., Tran, S., Toups Dugas, P.O., 2020b. A framework of input devices to support designing composite wearable computers. Lecture Notes in Computer Science 12182, 401–427. https://doi.org/10.1007/978-3-030-49062-1_28.

Khalaf, A.S., Alharthi, S.A., Tran, S., Dolgov, I., Toups Dugas, P.O., 2019. A taxonomy for selecting wearable input devices for mixed reality. Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces. Association for Computing Machinery, New York, NY, USA, pp. 403–408. https://doi.org/10.1145/3343055.3360759.

Khan, M.N.H., Neustaedter, C., 2019. An exploratory study of the use of drones for assisting firefighters during emergency situations. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 272:1–272:14. https://doi.org/10.1145/3290605.3300502.

Kira, Z., Potter, M., 2009. Exerting human control over decentralized robot swarms. Autonomous Robots and Agents, 2009. ICARA 2009. 4th International Conference on, pp. 566–571. https://doi.org/10.1109/ICARA.2000.4803934.

Kolling, A., Nunnally, S., Lewis, M., 2012. Towards human control of robot swarms. Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction. ACM, New York, NY, USA, pp. 89–96. https://doi.org/10.1145/2157689.2157704.

Kozak, M.N., Marsh, A.A., Wegner, D.M., 2006. What do i think you're doing? action identification and mind attribution. Journal of Personality and Social Psychology 90 (4), 543.

Kurebayashi, K., Hoffman, L., Ryan, C.S., Murayama, A., 2012. Japanese and american perceptions of group entitativity and autonomy: A multilevel analysis. Journal of Cross-Cultural Psychology 43 (2), 349–364. https://doi.org/10.1177/0022022110388566.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lee, M.K., Kiesler, S., Forlizzi, J., 2010. Receptionist or information kiosk: how do people talk with a robot? Proceedings of the 2010 ACM conference on Computer supported cooperative work. ACM, pp. 31–40.

Lee, S.-l., Lau, I.Y.-m., 2011. Hitting a robot vs. hitting a human: is it the same? Proceedings of the 6th international conference on Human-robot interaction. ACM, pp. 187–188.

Lickel, B., Hamilton, D.L., Sherman, S.J., 2001a. Elements of a lay theory of groups: Types of groups, relational styles, and the perception of group entitativity. Personality and Social Psychology Review 5 (2), 129–140.

Lickel, B., Hamilton, D.L., Sherman, S.J., 2001b. Elements of a lay theory of groups: Types of groups, relational styles, and the perception of group entitativity. Personality and Social Psychology Review 5 (2), 129–140.

Liu, M., Nancel, M., Vogel, D., 2015. Gunslinger: Subtle arms-down mid-air interaction. Proceedings of the 28th Annual ACM Symposium on User Interface Software &#38; Technology. ACM, New York, NY, USA, pp. 63–71. https://doi.org/10.1145/2807442.2807489.

Mann, S., 1997. Wearable computing: A first step toward personal imaging. Computer 30 (2), 25–32.

Marcoulides, G.A., 1989. Measuring computer anxiety: The computer anxiety scale. Educational and Psychological Measurement 49 (3), 733–739.

McDermott, D., Ghallab, M., Howe, A., Knoblock, C., Ram, A., Veloso, M., Weld, D., Wilkins, D., 1998. Pddl–the planning domain definition language–version 1.2. Yale Center for Computational Vision and Control, Tech. Rep. CVC TR-98-003/DCS TR-1165.

McRae, J.N., Gay, C.J., Nielsen, B.M., Hunt, A.P., 2019. Using an unmanned aircraft system (drone) to conduct a complex high altitude search and rescue operation: a case study. Wilderness & Environmental Medicine 30 (3), 287–290.

Meuter, M.L., Ostrom, A.L., Bitner, M.J., Roundtree, R., 2003. The influence of technology anxiety on consumer use and experiences with self-service technologies. Journal of Business Research 56 (11), 899–906.

Micire, M., Desai, M., Courtemanche, A., Tsui, K.M., Yanco, H.A., 2009. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces. ACM, New York, NY, USA, pp. 41–48. https://doi.org/10.1145/1731903.1731912.

Modares, J., Ghanei, F., Mastronarde, N., Dantu, K., 2017. Ub-anc planner: Energy efficient coverage path planning with multiple drones. 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 6182–6189.

Nomura, T., Suzuki, T., Kanda, T., Kato, K., 2006. Measurement of anxiety toward robots. ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp. 372–377.

Norman, D., 2013. The Design of Everyday Things: Revised and Expanded Edition, revised 2013. Basic Books.

PancockJr., C., Bass, E.J., Smith, P.J., 2016. UAS human factors and control station design standards: UAS control station literature review. Technical Report.

Penders, J., Alboul, L., Witkowski, U., Naghsh, A., Saez-Pons, J., Herbrechtsmeier, S., El-Habbal, M., 2011. A robot swarm assisting a human fire-fighter. Advanced Robotics 25 (1-2), 93–117. https://doi.org/10.1163/016918610X538507.

Ramchurn, S., Wu, F., Jiang, W., Fischer, J., Reece, S., Roberts, S., Rodden, T., Greenhalgh, C., Jennings, N., 2016. Human–agent collaboration for disaster response. Autonomous Agents and Multi-Agent Systems 30 (1), 82–111. https://doi.org/10.1007/s10458-015-9286-4.

Remes, B., Hensen, D., van Tienen, F., De Wagter, C., van der Horst, E., de Croon, G., 2013. Paparazzi: How to make a swarm of Parrot AR drones fly autonomously based on GPS. International Micro Air Vehicle Conference and Flight Competition (IMAV2013), pp. 17–20.

Robertson, J., 2018. Robot reincarnation: Rubbish, artefacts, and mortuary rituals. In: Cwiertka, K.J., Machotka, E. (Eds.), Consuming Life in Post-Bubble Japan: A Transdisciplinary Perspective. Amsterdam University Press, Amsterdam, Netherlands, pp. 153–174.

Rydell, R.J., Hugenberg, K., Ray, D., Mackie, D.M., 2007. Implicit theories about groups and stereotyping: The role of group entitativity. Personality and Social Psychology Bulletin 33 (4), 549–558. https://doi.org/10.1177/0146167206296956.PMID: 17363758

Rydell, R.J., McConnell, A.R., 2005. Perceptions of entitativity and attitude change. Personality and Social Psychology Bulletin 31 (1), 99–110. https://doi.org/10.1177/0146167204271316.PMID: 15574665

Salen, K., Zimmerman, E., 2004. Rules of Play: Game Design Fundamentals. MIT Press, Cambridge, MA, USA.

Sanchez-Lopez, J.L., Pestana, J., de la Puente, P., Suarez-Fernandez, R., Campoy, P., 2014. A system for the design and development of vision-based multi-robot quadrotor swarms. 2014 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, pp. 640–648.

dos Santos, T.D., Santana, V.F.d., 2018. Computer anxiety and interaction: a systematic review. Proceedings of the Internet of Accessible Things, pp. 1–10.

Seppanen, H., Mäkelä, J., Luokkala, P., Virrantaus, K., 2013. Developing shared situational awareness for emergency management. Safety Science 55, 1–9. https://doi.org/10.1016/j.ssci.2012.12.009.

Smith, E.R., Mackie, D.M., 2008. Intergroup emotions. Handbook of emotions 3, 428–439.

Smith, E.R., Seger, C.R., Mackie, D.M., 2007. Can emotions be truly group level? evidence regarding four conceptual criteria. Journal of Personality and Social Psychology 93 (3), 431–446.

Smith, E.R., Sherrin, S., Fraune, M.R., Šabanović, S., 2020. Positive emotions, more than anxiety or other negative emotions, predict willingness to interact with robots. Personality and Social Psychology Bulletin.0146167219900439

Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R.W., Pentland, A., 1997. Augmented reality through wearable computing. Presence:

Teleoper. Virtual Environ. 6 (4), 386–398. https://doi.org/10.1162/pres.1997.6.4.386.

Systems, T., 2016. Trx develops proximity system to increase firefighter safety. [Online; posted 21-April-2016].

TAP, 2016. Welcome to TAP. http://tapwithus.com/2016/05/09/welcome-to-tap/.

Toups Dugas, P.O., Hamilton, W.A., Alharthi, S.A., 2016. Playing at planning: Game design patterns from disaster response practice. Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. ACM, New York, NY, USA, pp. 362–375. https://doi.org/10.1145/2967934.2968089.

Toups Dugas, P.O., Kerne, A., 2007. Implicit coordination in firefighting practice: Design implications for teaching fire emergency responders. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 707–716. https://doi.org/10.1145/1240624.1240734.

Toups Dugas, P.O., Kerne, A., Hamilton, W.A., 2011. The team coordination game: Zero-fidelity simulation abstracted from fire emergency response practice. ACM Trans. Comput.-Hum. Interact. 18 (4), 23:1–23:37. https://doi.org/10.1145/2063231.2063237.

Vergouw, B., Nagel, H., Bondt, G., Custers, B., 2016. Drone technology: Types, payloads, applications, frequency spectrum issues and future developments. The Future of Drone Use. Springer, pp. 21–45.

Waharte, S., Trigoni, N., 2010. Supporting search and rescue operations with uavs. 2010 International Conference on Emerging Security Technologies, pp. 142–147. https://doi.org/10.1109/EST.2010.31.

Waytz, A., Gray, K., Epley, N., Wegner, D.M., 2010. Causes and consequences of mind perception. Trends in cognitive sciences 14 (8), 383–388.

Weichert, F., Bachmann, D., Rudak, B., Fisseler, D., 2013. Analysis of the accuracy and robustness of the leap motion controller. Sensors 13 (5), 6380–6393. https://doi.org/10.3390/s130506380.

Weldon, W.T., Hupy, J., 2020. Investigating methods for integrating unmanned aerial systems in search and rescue operations. Drones 4 (3), 38.

Westman, M., Eden, D., 1996. The inverted-u relationship between stress and performance: A field study. Work & Stress 10 (2), 165–173.

Wullenkord, R., Fraune, M.R., Eyssel, F., Šabanović, S., 2016. Getting in touch: How imagined, actual, and physical contact affect evaluations of robots. 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE, pp. 980–985.

Yuki, M., 2003. Intergroup Comparison versus Intragroup Relationships: A Cross-Cultural Examination of Social Identity Theory in North American and East Asian Cultural Contexts. Social Psychology Quarterly 66 (2), 166–183. https://doi.org/10.2307/1519846.

**Dr. Marlena R. Fraune** is an Assistant Professor leading the Intergroup Human-Robot Interaction Lab in the Department of Psychology at New Mexico State University. She received her B.A. in Psychology at Beloit College in 2013 and her PhD at Indiana University in 2018. Dr. Fraune studies how to make human-robot more efficient and enjoyable. She has done so nationally and abroad as a visiting researcher at Toyohashi University of Technology (2014, 2016), Bielefeld University (2015), and at the United States Air Force Research Lab (2017).

**Ahmed S. Khalaf**, MA, is currently a PhD student in Computer Science at New Mexico State University, working under the supervision of Dr. Phoebe O. Toups Dugas His research focuses on developing wearable interface design to direct agent teams with adaptive autonomy. He also works as a Lecturer at Al-Baha University, Saudi Arabia.

**Mahlet Zemedie** is currently a masters' student in Computer Science at New Mexico State University. She received her bachelor's in Electrical Engineering in 2016. She has been working as a Support Engineer at Ethio Telecom, Addis Ababa, Ethiopia. She is currently working as a Research Assistant under the supervision of Dr. Marlena R. Fraune and Dr. Z Toups.

**Dr. Bill Hamilton** is an Assistant Professor of Computer Science at New Mexico State University. He leads the Participatory Live Experiences (PLEX) Lab. Dr. Hamilton earned his B.S. and Ph.D. in Computer Science from Texas A&M University. Dr. Hamilton's work investigates online communities and teams and how to support participation through the design of digital media. He has investigated these issues in the contexts of online video game streaming, online learning, and political discourse.

**Poom Pianpak** is currently a PhD student in Computer Science at New Mexico State University, working under the supervision of Dr. Son Tran. His research is in Artificial Intelligence focusing on knowledge representation and reasoning, multi-agent systems, and multi-agent path finding.

**Dr. Son Tran** received his Diplom Mathematiker degree in Mathematic Cybernetics and Comuting Technique in 1986 from the Technical University of Dresden, Germany, and his Master's and Ph.D. degrees in computer science from the Asian Institute of Technology, Thailand, and University of Texas at El Paso in 1993 and 2000, respectively. Before joining the Computer Science Department at New Mexico State University as an Assistant Professor in 2001, he was a post-doctoral at the Knowledge Systems Laboratory, Computer Science, Stanford University.

**Zahra Naminimianji** is a software engineer currently working at Don't Panic Labs, in Lincoln, Nebraska. She received her M. Sc. degree from the Department of Computer Science at NMSU in 2019. During her graduate studies, she worked on developing an architecture for simulating drones in a mixed reality game to examine use of them in search and rescue scenarios, in PIxL Lab under supervision of Dr. Z Toups.

**Dr. Phoebe O. Toups Dugas** is an Associate Professor of Computer Science at New Mexico State University, where they lead the Play and Interactive Experiences for Learning Lab. Dr. Toups earned their B.A. in Computer Science from Southwestern University in 2003 and their Ph.D. in Computer Science from Texas A&M University in 2010. They design systems at the intersection of games, collaboration, wearables, and disaster response and study the design of games using qualitative methods.

**Dr.Sultan Alharthi** is an HCI and Games researcher. He received his PhD in Computer Science from New Mexico State University. His research explores novel approaches to support collective sensemaking, planning, team awareness, and collaboration in distributed environments.