User-defined Swarm Robot Control

Lawrence H. Kim
Stanford University
Stanford, CA
lawkim@stanford.edu

Daniel S. Drew
Stanford University
Stanford, CA
dsdrew@stanford.edu

Veronika Domova
ABB CRC
Vasteras, Sweden
verdo29@liu.se

Sean Follmer
Stanford University
Stanford, CA
sfollmer@stanford.edu

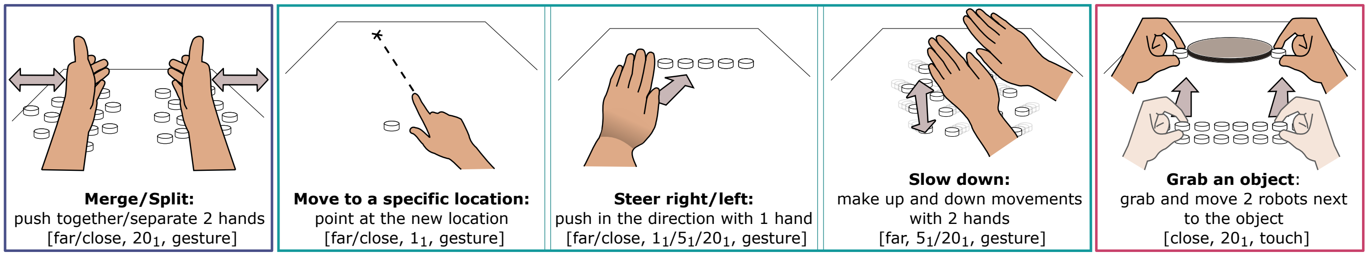

Figure 1. We ran an elicitation study to better understand how users control a swarm of robots. Examples of high agreement interactions are shown here. People used varying number of fingers/hands and different interaction modalities such as gesture and touch. The first two values inside the brackets indicate the proximity and number of robots for the interaction and the last value indicates the interaction modality. The colored boxes indicate the task type that it belongs to. Blue, teal, and red boxes represent inter-robot interaction, navigation, and object manipulation task types.

ABSTRACT
A swarm of robots can accomplish more than the sum of its parts, and swarm systems will soon see increased use in applications ranging from tangible interfaces to search and rescue teams. However, effective human control of robot swarms has been shown to be demonstrably more difficult than controlling a single robot, and swarm-specific interactions methodologies are relatively underexplored. As we envision even non-expert users will have more daily in-person encounters with different numbers of robots in the future, we present a user-defined set of control interactions for tabletop swarm robots derived from an elicitation study. We investigated the effects of number of robots and proximity on the user's interaction and found significant effects. For instance, participants varied between using 1-2 fingers, one hand, and both hands depending on the group size. We also provide general design guidelines such as preferred interaction modality, common strategies, and a high-agreement interaction set.

Author Keywords
Swarm Robot Control; Multi-robot Control; Swarm Robotics; Swarm User Interface; Elicitation study

CCS Concepts
•**Human-centered computing** → **User studies;** *Interaction techniques; User centered design;*

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00
DOI: https://doi.org/10.1145/3313831.3376814

INTRODUCTION
Robots are increasingly being deployed across personal, commercial, and industrial sectors, with application spaces ranging from elderly-care social assistants to members of firefighting teams. We are moving towards a society where humans are actually outnumbered by autonomous and semi-autonomous agents in both their home and work lives, similar to the vision of "ubiquitous robotic interfaces" described in [20]. Some of these robots will work together in small groups, typically thought of as "multi-agent systems." For applications where things like areal distribution, low unit cost, and robustness to agent failure are critical, research has begun towards the development of swarm systems, where, as in the natural world of insects, large (>10) groups of robots must work together to become more than the sum of their parts [38]. This emerging field of swarm robotics presents many challenges in the area of human-swarm interaction (HSI), including the cognitive complexity of solving tasks with swarm systems, state estimation and visualization, and human control of the swarm [23].

While HSI researchers have developed numerous ways to control a swarm of robots in situ [15, 3, 37], they all share one important limitation: lack of consideration for user's preferences and intuition. Instead of integrating sensors that can sense a set of user-defined interaction, prior work has mostly focused on finding a set of interactions that the existing robotic sensors can detect, and then semi-arbitrarily mapping these interactions to a set of control commands. This is a problem, as such interaction vocabularies may only be effective for domain experts or the designers themselves and could present a steep learning curve for novice users. As we see a near future with wider adoption of swarm robot technologies that will constantly exist in both our public and private environments, we focus on proximal control that could function on an encountered basis even for novice users. Thus, we sought to ground HSI through user-centric approaches. While prior works have studied interaction with a single robot and ran elicitation studies on control of a team of drones [34, 17], it is unclear how those results map to grounded, large agent count multi-robot systems.

To better understand how users *prefer* to interact with swarm robots, we present an elicitation study with up to 20 centimeter-scale tabletop robots. As prior work has shown that number of robots and proximity to the robots affect human's perception and behavior [47, 30], we also investigated the effects of these variables on user's desired input method. The tasks ranged a large span of possible interactions, including formation control, parameter setting, cooperative manipulation, and human-robot teaming concepts (e.g. "follow me"). Care was taken to abstract implementation details from the users in order to elicit multi-modal input schemes which include gestures, touch and verbal interactions. Using the study results, we compiled a user-defined interaction set with interactions based on not only referents but also number of robots and proximity. We also examine overall trends on interaction modality, taxonometric breakdown, and agreement scores to better understand how participant interact. These results can inform the design and sensing required to support rich interaction with swarm robots.

RELATED WORK

The most relevant related areas of research to this work include swarm robotics, studies of control strategies for human operators of multi-agent systems, and prior elicitation-based studies of natural human-robot interaction methods.

Swarm Robotics

Swarm robotics is a field concerned with the coordination of very large (greater or much greater than ten agent) groups of relatively simple, and often small, robots [38]. A survey by Dudek et al. [10] established taxonomies for swarm system implementations (e.g., based on communication topology or agent processing power) as well as their envisioned tasks (e.g., inherently multi agent, inherently single agent) which have been used to guide numerous later investigations.

Another survey by Kolling et al. [23] describes and categorizes common swarm control strategies found in the literature; the choice between these common control strategies is often dictated by a system's described level-of-automation (LOA) [41], an idea which has been extended to include variable control schemes depending on a desired LOA on a per-task basis [8]. In an effort to uncover more natural interaction modes we have avoided describing to the users either the system implementation or context of the tasks in enough detail for them to rigidly fit into these taxonomies, although doing so has precluded us from assessing variable user-desired interaction schemes according to an autonomy spectrum.

While other researchers have investigated the manipulation of swarm implementation parameters (e.g., group size, agent speed and coordination) on human physiological response [36] and affect [20], we instead seek to find how parameters like group size and distance from the user change desired operator control methodology.

Swarm User Interfaces

Inspired by the dream of "tangible bits" and "programmable matter" [16], one envisioned use case for robot swarms are as physical mediators of digital data; that is, a tangible user interface (TUI). Researchers have begun to develop swarm-based tangible user interfaces for data visualization [24, 25], education [32], and social interaction [21]. More versatile shape-changing swarm robots have been demonstrated that could potentially span multiple application spaces [46]. Beyond using swarms as displays, the vision of ubiquitous robotics [20] builds on that of ubiquitous computing [50] to imagine versatile robotic interfaces that are mobile, can physically manipulate the world around them, can convey information through dynamic action, and can form tangible links between the user and the physical world. Typical studies of swarm interfaces, however, focus on their efficacy at performing a given set of tasks, whereas in this work we instead look towards finding natural methods for future users to command and interact with "ubiquitous" swarm systems.

Human Control of Multi-agent Systems

There is a good deal of research into effective computer interfaces for controlling swarms and multi-agent systems [22, 27], but here we focus on proximal control that could function on an encountered basis (i.e., without dedicated interface hardware like a mouse and GUI). Most of the prior literature in this style seeks to demonstrate full end-to-end implementations in order to prove the viability of things like gesture-based control. A variety of possible sensor suites have been used for this purpose, including user-wearable attitude [15] or EEG sensors [43], centralized fixed vision systems [3], multi-touch screen [18, 21], and consensus-based distributed vision algorithms [31]. The most relevant work specifically investigates proximal interactions with wheeled multi-agent/swarm systems with well-defined task sets such as in [37]. In our study, we narrowly focus on single-operator multi-modal interaction within an on-table environment. In contrast with tabletop swarm interface work like Reactile [45], which was explicitly a tangible interface made for only physical interactions, we let the users decide how they interact with the robots.

Human-Robot Interaction Elicitation Studies

In an elicitation study, participants are prompted to choose and demonstrate their own preferred input method (and specific action) for a given task. Although their efficacy in uncovering natural interaction schemes has been validated in other areas [14, 53, 29], elicitation studies in the context of human-swarm interaction remain rare.

There are some examples of elicitation studies for control of UAVs in the literature, but the increased low-level control complexity for safely operating high numbers of proximal drones means that the number of robots interacted with in these studies is typically limited. A multimodal (gesture, sound, and touch) elicitation has been performed with a real single UAV [1], gesture-only elicitation for up to four real UAVs at a time [34], and for a swarm of 10 UAVs with voice and gesture multimodal input in simulation [17]. In contrast, working with on-table wheeled robots lets us deploy, without computer-rendered or VR simulation, relatively numerous

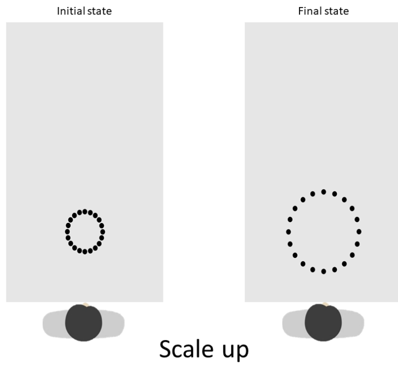

Figure 2. Sample pictorial prompt for the "Scale up" referent.

groups of small robots (i.e., closer to future envisioned swarm systems) operating on a 2-dimensional workspace. This capability provides us with the unique opportunity to investigate the effects of group size and robot proximity on user input preference across a wide swath of example tasks, without worry that our results will suffer from the documented "reality gap" that exists in simulated (as opposed to implemented in hardware) human-robot interaction studies [36].

ELICITATION STUDY ON IN SITU SWARM ROBOT CONTROL

To better understand how users prefer to interact with a swarm of robots, we conducted an elicitation study on swarm robot control. Our study results can inform what types of sensors are needed to enable fluid interaction between users and a swarm of robots. We have pre-registered our elicitation study at OSF.io (https://osf.io/r8fnc) and all raw data along with study results are freely available at https://osf.io/dkja9/ as well.

Hypotheses

In addition to understanding how users interact with a swarm of robots, we investigated the effects of a few key interaction parameters: number of robots and proximity to the robot(s).

H1: Number of robots will affect how users interact

Researchers have shown that the number of robots can significantly alter how people perceive the robots when viewing their motion [36] or being touched by them [21]. Researchers have also developed different ways to teleoperate or remotely control a swarm of agents such as leader-follower [7], selection and beacon control [22], and physicomimetics [42]. Thus, we hypothesize that users will also adapt their interaction method for in situ control based on the number of robots.

H2: Proximity to the robot(s) will affect how users interact

Literature in Human-Robot Interaction (HRI) has shown that humans perceive robots differently based on their proximity as well as prefer robots that exhibit proxemic behavior [47, 30]. Cauchard et al. have reported that when the robots were closer, users tended to use smaller motions [6]. Thus, we also hypothesize that proximity to the robot(s) will change how users choose to interact with a swarm of robots.

Methodology

We employed a similar method as in [6] with slight modifications to address the real-time controllability of the robots and to improve accessibility for non-native English speakers. Instead of a complete Wizard-of-Oz (WoZ) elicitation study, we conducted a semi-WoZ study due to the difficulty of controlling a large number of robots impromptu. We preprogrammed each referent and timed the initiation once the participants completed their interaction. As shown in Fig. 2, we displayed pictorial [1] instead of purely textual [6] prompts for the referents as they reduce verbal biasing and lower language barriers for non-native English speakers. These prompts include initial and final state of the robots, and the task title.

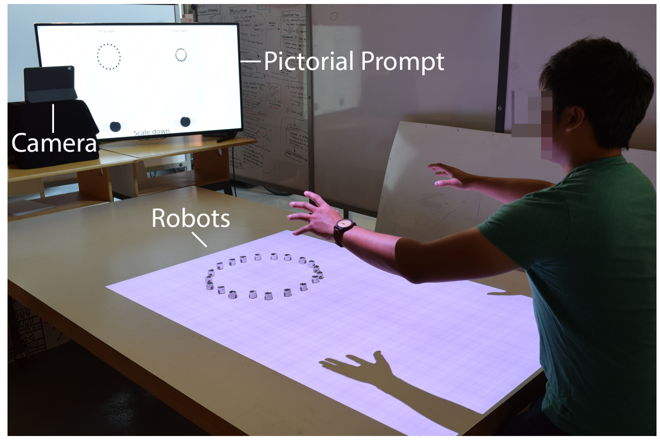

Figure 3. Setup for the control elicitation study: After being prompted through a television monitor, participants interact with 20 robots on a table while standing.

Apparatus

We used a modified version of Zooids, a wheeled multi-robot platform [24], with a higher gear ratio motor (136:1) in order to render more accurate movements similar to [20]. As shown in Fig. 3, the pictorial prompts for the referents were displayed on a 50 inch television monitor while a video camera was used to record participants' interaction. Depending on the referent, up to 20 robots moved within a 1.4 x 0.89 m work space (i.e. projected space) on a table.

Referents

To generate the list of referents for this study, we combined the control and interaction commands from prior literature in swarm or multi-robot control [3, 15, 17, 48, 45, 43, 31, 37, 40] and interaction with a single or multiple drones [6, 1, 13, 34, 4] as shown in Table 1. For referents under the "Robot Selection" and "Inter-Robot Interaction" categories as well as "Move here and there" and "Grab an object" referents, only 20 robots were used since these referents are most relevant when there are a significant number of robots. To reduce study duration and user fatigue, we combined pairs of referents from prior works that were similar and opposite of each other such as "move closer/away", "steer left/right", and "move faster/slower". For these pairs, each pair instead of each referent was presented under all 6 conditions (3 (# of robots) x 2 (proximity)).

Participants

15 participants were recruited (7 M, 8 F) from our institution. Age ranged from 19 to 41 (average: 29, std: 5.7). Their edu-

Category | subcategory		Referents [related work]
Robot Selection (20 robots)		Select one robot [15, 31, 3, 4, 34, 43, 37] Select a group of robots [15, 31, 3, 4, 34, 43, 37] Select all robots [15, 31, 3, 4, 34, 43, 37]
Inter-Robot Interaction (20 robots)		Form a circle [43, 34, 45] Split/merge [37, 17] Scale up/down [3, 34, 45, 13] Rotate [3, 13]
User-Robot Interaction (1, 5, 20 robots)	Navigation	Move closer/away [6, 1] Follow me [6, 1]
	ETC	Get attention [6, 1]
Robot-Environment Interaction (1, 5, 20 robots)	Navigation	Move to a location [15, 17, 13] Move here and there (only with 20 robots) Steer left/right [37, 43, 17] Stop [37, 43, 6, 1] Move faster/slower [15] Follow trajectory [3]
	Manipulation	Grab an object (only with 20 robots) [48] Push an object [40]

Table 1. List of referents used in the elicitation study.

cational backgrounds ranged from engineering (9), computer science (2), and others (4). They were compensated $15.

Procedure

For each referent displayed on the screen, participants were instructed to perform any interaction method that they choose to complete the given task. They were told to focus on how they would prefer to interact as opposed to focusing on whether the robot(s) could understand their interaction. No suggestions were given on how to interact with the robot(s). After the participants completed each interaction, they explained their interaction in 1-2 sentences and rated their interaction on a 7-point Likert scale in terms of suitability (i.e. how well their interaction fit the task), simplicity (i.e. how simple their interaction was), and precision (i.e. how precise the interaction was). To become familiar with the process, we included 3 practice trials including one basic referent (move closer) and two referents that pilot subjects found more complex (follow me, steer right) in the beginning. They then proceeded to the actual experiment with 76 conditions in randomized order. After the participants completed the entire study, they filled out a post-test survey and had a short interview regarding their experience.

ANALYSIS

Taxonomy

To understand what types of gesture, touch, and verbal interactions were used, we analyzed them using a modified version of the existing taxonomies in surface gesture [53], manipulation [5], and illocutionary acts [39] as shown in Table 2.

GESTURE		
# of finger /hands	*1-2 fingers* *one hand* *both hands*	1-2 fingers are used. more than two fingers used. Both hands are used.
Form	*static pose* *dynamic pose* *static pose & path* *dynamic pose & path*	Hand pose is held in one location. Hand pose changes in one location. Hand pose is held as hand moves. Hand pose changes as hand moves.
Nature	*deictic* *symbolic* *physical* *metaphoric* *abstract* *iconic*	Indicative or pointing gesture. Gesture visually depicts a symbol. Gesture acts physically on objects. Gesture indicates a metaphor. Gesture-referent mapping is arbitrary. Gesture depicts aspects of spatial images, action, people or objects.
Binding	*robot-centric* *user-centric* *world-dependent* *world-independent*	Location defined w.r.t. robots. Location defined w.r.t. user. Location defined w.r.t. world. Location can ignore world features.
Flow	*discrete* *continuous*	Response occurs after the user acts. Response occurs while the user acts.
TOUCH		
# of finger /hands	*1-2 fingers* *one hand* *both hands*	1-2 fingers are used. More than two fingers used. Both hands are used.
# of robots touched	*one robot* *few robots* *many robots*	User manipulates one robot. User manipulates 2-4 robots. User manipulates more than 4 robots.
Control paradigm	*leader-follower* *follow crowd* *control all*	User manipulates only one robot. User manipulates a subset of robots. User manipulates all of the robots.
Manipulation type	*NP, NM* *NP, M, NW* *P, NM* *P, M, NW*	Non-prehensile with no motion. NP with motion but not within hand. Prehensile with no motion. P with motion but not within hand.
VERBAL		
Illocutionary acts	*Directives* *Expressives*	Get the hearer to do something. Express attitudes and emotions.

Table 2. Taxonomy of Gesture, Touch, and Verbal Interactions.

Gesture

For gesture, we labelled each interaction by the number of fingers/hands used (one or two fingers, one hand, both hands) and by the four dimensions (form, nature, binding, flow) from the taxonomy of surface gesture [53]. For the four dimensions of surface gesture [53], we modified categories within each dimension to better fit our context. For the form dimension, we removed "one-point touch" and "one-point path" as our interaction space is not limited to a 2-D space. Instead, we added "deictic" in the nature dimension as well as "iconic" to better classify 3-D gestures. For the binding dimension, we removed "mixed-dependencies" as we didn't observe any corresponding interaction and added "user-centric" to better accommodate user-robot interactions.

Touch

For touch interactions, we classified each interaction by the number of fingers/hand used (one or two fingers, one hand, both hands), number of robots touched (one, few (2-4), many), number of robots touched simultaneously (one, few, many), control paradigm (leader-follower, follow crowd, control all) as well as using the "contact" part of the human manipulation taxonomy [5]. As we observed no within hand manipulation either non-prehensile or prehensile (NP, M, W or P, M, W), we excluded those categories.

Verbal

Searle classified illocutionary acts into five basic categories: representatives, directives, commissives, expressives, and declarations [39]. However, as we only observed directives and expressives during the study, we labelled each verbal interaction as one or the other.

Reliability

Two of the authors coded the interaction based on the recorded videos. To improve agreement, all four authors discussed the coding scheme and coded some common interactions together. To measure the Inter-Rater Reliability (IRR), the two coders independently coded 15 conditions from two different participants and calculated the unweighted Cohen's Kappa for 11 items (average = 0.79, std = .17). For the remaining conditions, only one rater coded each condition.

Agreement Score

After grouping identical interactions within each referent, we computed the agreement score [52] for each referent, number of robots, and proximity.

Statistical Analysis

We used Fisher's exact test of independence to test whether the proportions of one nominal variable are different depending on the value of another nominal variable as the sample size is relatively small (n<1000) [26]. Then, Bonferroni-corrected post-hoc tests were used to determine which pairs are significantly different. For instance, Fisher's test was used to test whether the proportions of number of hands used for gesture interactions were different based on the number of robots. To compare the means of the participant's ratings on their interaction for different number of robots, proximity, and tasks, we used N-way ANOVA followed by Bonferroni-corrected post-hoc tests.

RESULTS & DISCUSSIONS

Overall Trends

Interaction Modality

We categorized each interaction into one of the following interaction modalities: gesture, touch, verbal commands, and combinations of them. Figure 4 presents the breakdown of interaction modalities used across all conditions. For multi-modal interactions, they are counted in all relevant categories. For example, interactions with both gesture and verbal commands are counted in "Gesture", "Verbal", and "G+V". When looking at these overall results in the context of prior work, we see some similar trends to single robot interaction across different types of robots in terms of interaction modality - for

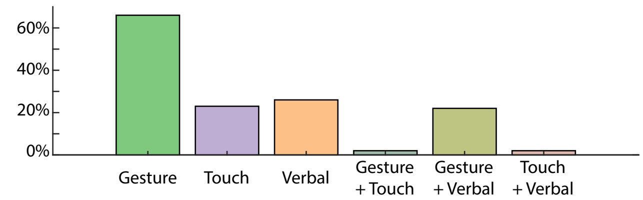

Figure 4. Breakdown of Interaction Modality for all conditions

example, our results for cm-scale wheeled robots are similar to the results found by Abtahi et al. for a single caged "safe" aerial drone [1]. Yet, our results are quite different than that of uncaged "unsafe" drones, potentially due to the non hazardous nature of our small wheeled robots. However, our results are less directly comparable to other studies which did not explore the use of touch or direct manipulation for control of many robots as the study is done in a virtual environment [17, 34]. Yet, similar to [17], we also observed that the majority of the speech commands were accompanied by a gesture.

Taxonometric Breakdown

Using the taxonomies in Table 2, we labelled each interaction and the taxonometric breakdown is shown in Figure 5.

Gesture:

The majority of the gestures had static pose and path form. In terms of the nature, there is heavy reliance on the use of physical, symbolic, and deictic gestures. This suggests that similar to how *physics engine* is used for surface recognition [53, 51], swarm robot control could also benefit from a physics-based detection algorithm. In addition, it is important for the recognition algorithm to know common symbolic gestures as participants expected the robots to understand common symbolic gestures such as a "stop" gesture with palm showing toward the robot(s) or a "calming" gesture for the slow down referent with hands moving down slowly.

Most gestures were defined with respect to the robots and almost 90% were discrete. The flow was most likely influenced by two factors. First, many of the referents such as robot selection tasks and get attention task are simple with no intermediate steps. Thus, there was no need for continuous input. Second, the robots were not fully controlled impromptu but rather had pre-programmed behaviors with investigator-timed initiation. This setup did not allow any adjustments to the robots' behavior after the initiation and thus discouraged participants from using continuous interactions.

Touch:

55% of touch interactions involved one or two finger touch to physically manipulate one robot. When the task involved more robots, participants relied on different control paradigms such as leader-follower (where they only manipulate one robot and expect the rest to follow), and follow crowd (where they manipulate a subset of the entire group and expect the rest to follow) as it was difficult to grab and manipulate all of the robots at the same time.

We also observed that participants tended to use other modalities when more robots were involved. For instance, P2 wrote in the survey that "...I wasn't sure how to grab them all so it led

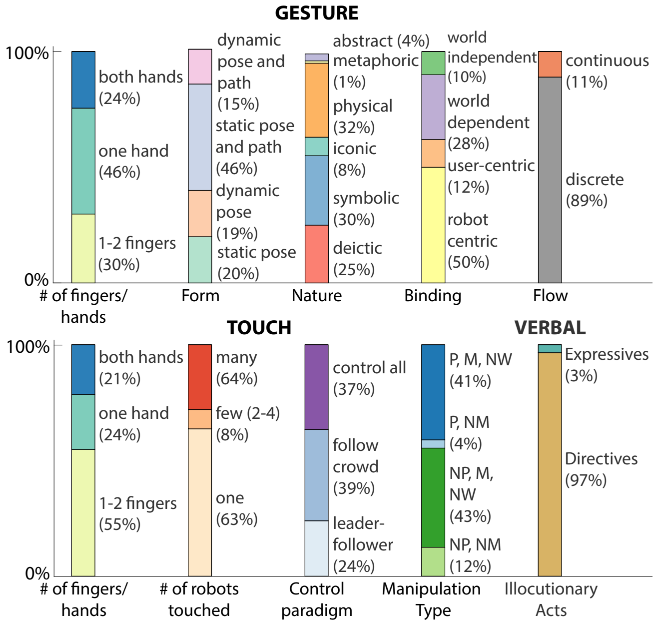

Figure 5. Taxonometric breakdown for Gesture, Touch, and Verbal Interactions across all conditions

me to think of other ways to direct them other than touching them." while P8 mentioned that "more robots there were, the more I was inclined to give a global input. Like audio."

43% of the touch interactions were non-prehensile with motion not within hand (C,NP,M,NW) while 41% were prehensile with motion not within hand (C,P,M,NW). We saw very little counts of prehensile manipulation with no motion (P, NM) as participants usually grabbed the robots to move them somewhere. Even for tasks where the robots do not need to move such as robot selection tasks, most interaction involved tapping or touching the robot(s) while there were only few instances of pure grasp with no motion.

Verbal:
97% of the verbal interactions were directives (i.e., commands). However, there were a few cases where the participants used expressives instead to imply what the robots should do. For instance, a participant said "you guys are moving too fast" to imply that the robots should move slower, whereas another said "you guys are too tight" for the "scale up" referent. This suggests that some users may not always explicitly communicate the desired action and that a voice recognition algorithm will need to infer the user's intention.

Agreement
The agreement scores across all interaction modalities for each referent, number of robots, and proximity are shown in Figure 9. The overall average agreement scores for gesture, touch, and verbal interactions independently are $A_G = 0.3$, $A_T = 0.56$, and $A_V = 0.37$.

User-Defined Interaction Set

The user-defined interaction set was generated by taking the most frequent interaction for each referent. If the same interaction was used for different referents thus creating conflict, the interaction was assigned to the referent with the largest group. The resulting interaction set is shown in Figures 1, 6, and 7. For each interaction, we describe the proximities, numbers of robots that the interaction is representative of, and the interaction modality of the interaction. These are represented by the three values inside the bracket after the description of the interaction. The subscript 1 or 2 under "far" or "20" indicates that the interaction is the first or second most frequent interaction for the "far" or "20" robot condition. For example, for the top-left interaction "draw a circle with a finger", it is the most frequent gesture interaction for both far and close proximity condition. Different task categories are indicated by the colored box around the illustration. Blue, dark green, orange, red, teal, and maroon boxes represent represent inter-robot interaction, robot selection, user-centered navigation, getting attention, navigation in environment, and object manipulation task types.

Prior work has shown aliasing significantly improves the input guessability [12, 52]. In our interaction set, five referents are assigned 1 interaction, eleven referents have 2 interactions, seven referents have 3 interactions, and one referent has 5 interactions. Out of the 53 interactions in the final set, 15 (28%) are performed with one or two fingers, 21 (40%) are performed with one hand, and 17 (32%) with both hands.

Effects of Number of Robots

As we hypothesized, the number of robots had significant effects on the participant's interaction. It had a positive correlation with the number of hands used, affected the control paradigm they used for touch interactions as shown in Figure 8d, and had a negative correlation with participant's simplicity ratings of the interaction ($p < 0.05$).

Number of Fingers/Hands:
The number of robots had a significant effect on how many hands the participants chose to use. When interacting with more robots, participants increased the number of their hands for both their gesture and touch interactions (both $p < 0.001$) as shown in Figure 8a and 8b. To control a single robot, they used one/two fingers or a single hand whereas they relied on both hands to interact with 20 robots. The post-test survey revealed that participants were indeed mindful of how they use the number of hands. For example, P9 wrote "If there was one robot I was more likely to use one finger, versus all of the robots, I wanted to ... use two hands.", while P5 wrote "I often wanted to use both hands when interacting with a group of robots, even though I knew already a single hand could work the same." As P5 mentioned, even though there was no need or instruction to use more hands for more robots, participants felt the need to use both hands as confirmed by the study results. Although not explicitly studied, this trend is hinted in the interaction set from [34].

In addition to using both hands, we also observed that participants tried to match the size of their hands, via spreading fingers, to the number of robots. For instance, P15 wrote in the post-test survey "I tried to spread my hands wide enough to cover the whole area of the robots" while P4 mentioned that "I tended to use all my fingers with larger groups." Further investigation will be needed to confirm this.

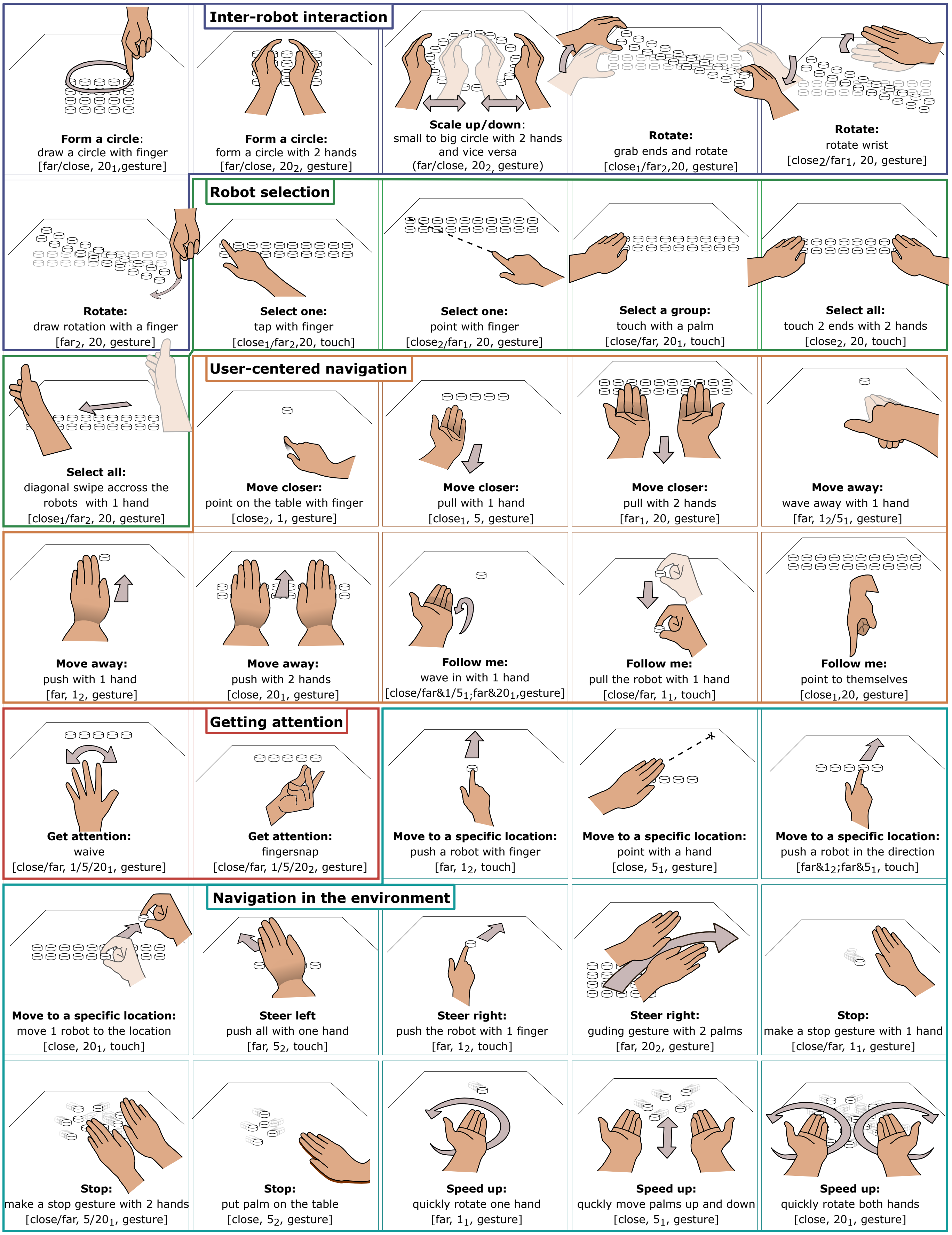

Figure 6. User-Defined Interaction Set. To save space, reversible gestures (split/merge, scale up/down, steer left/right) have been combined and the interactions shown on the first page are not shown here.

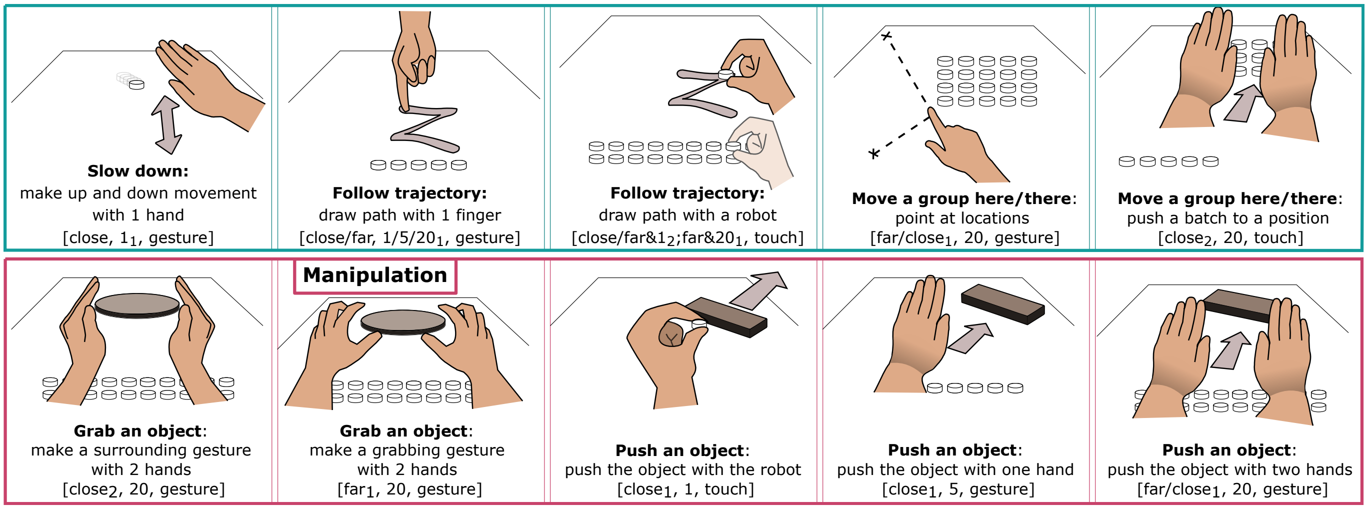

Figure 7. User-Defined Interaction Set

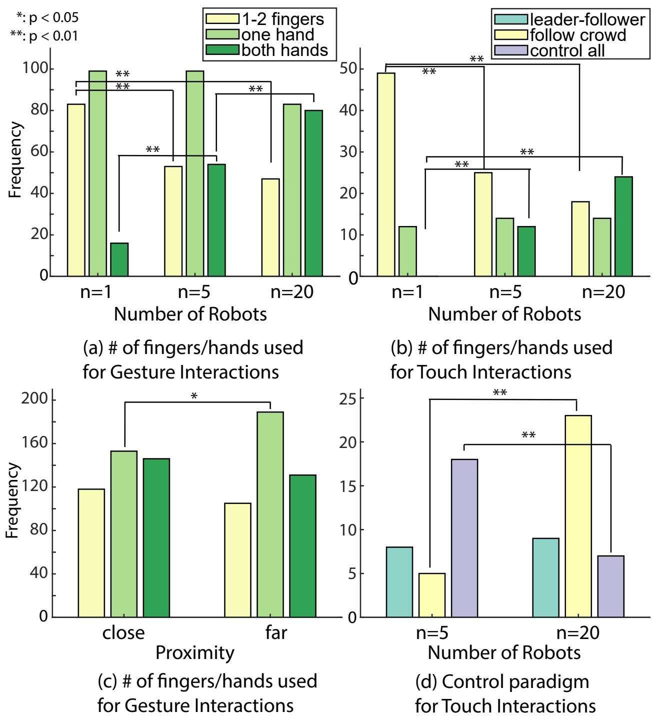

Figure 8. Number of robots have significant impact on the number of fingers/hands used for (a) gesture and (b) touch interactions. (c) For gesture interactions, proximity significantly affects the number of fingers/hands used. (d) For touch interactions, the number of robots has significant effects on the control paradigm.

While our taxonomy does not capture the magnitude of the gesture (i.e., how big spatially the gesture is), participants also mentioned that they used "bigger gesture[s] for larger number of robots" (P3), made "larger motions when there were more robots" (P14).

Control Paradigm for Touch Interactions:
When interacting with many robots, participants were less likely to directly manipulate all of the robots as shown in Figure 8b ($p < 0.001$). To overcome this, they either used a leader-follower or a follow crowd control paradigm, where they directly manipulate either just one or a subset of the robots respectively. We see this change as the number of robots increases from 5 to 20, as shown in Figure 8b.

Simplicity Ratings:
The number of robots significantly affected the participant's simplicity ratings on their interaction ($p < 0.05$). The simplicity ratings for interactions with one robot were higher than those of interactions with 20 robots.

Effects of Proximity

As hypothesized, proximity to the robots had significant effects on how participants chose to interact in terms of number of hands used as shown in Figure 8c and their self-reported ratings on how precise their interaction was ($p < 0.05$).

Number of Finger/Hands:
Proximity had significant effect ($p < 0.05$) on the number of hands used for gesture as shown in Figure 8c. When the robot were far away, participants used one hand more often than when the robots were close. One potential reason for this is that when the robots were far away, we found participants tended to lean forward over the table to make the gesture clearer to the robots; it may have become more convenient or stable for the users to use only one hand in such a position.

Precision Ratings:
The proximity to the robots significantly affected the participant's precision ratings on their interaction ($p < 0.05$). The precision ratings for close proximity conditions were higher than those for far proximity conditions.

Trends within Each Referent Category

For each referent category, we compared its data with that of the remaining referents. For instance, for the robot selection category, we compared its data with that of referents not in the robot selection category.

Robot Selection
For robot selection tasks, participants relied significantly more on touch interactions than for non-selection tasks ($p < 0.01$).

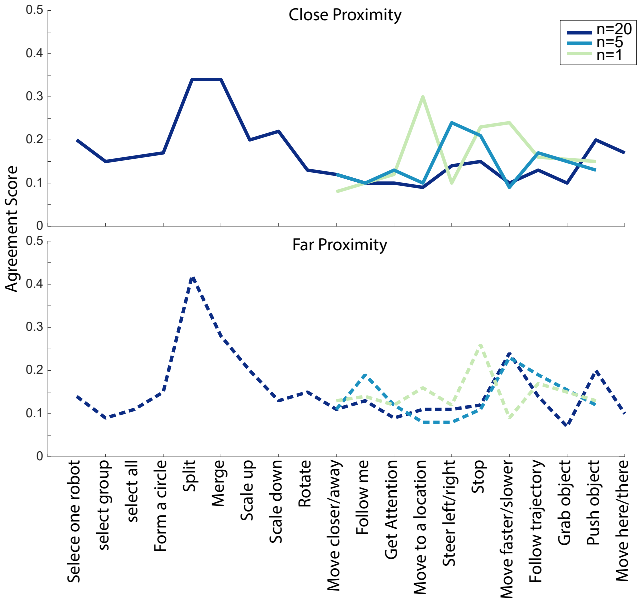

Figure 9. Agreements for each referent across all interaction modalities.

Also, proximity had a significant effect on the interaction modality ($p < 0.05$). When selecting robot(s) in close proximity, participants tended to use touch interaction more frequently than when selecting remote robot(s). As shown in Figure 10, participants used significantly fewer two-handed ($p < 0.01$) and "static pose and path" form gestures while using more "static pose" form gestures ($p < 0.01$) compared to the non-selection tasks. The nature of the gestures was also different as there was a significant increase ($p < 0.001$) in use of deictic gestures and decreases in physical and abstract gestures. Almost all of the gestures were discrete and robots-centric. These results could inform the design of interaction techniques for selection tasks with many robots, which could be used in many applications such as infrastructure maintenance [35], search-and-rescue [44], data physicalization [25], and environmental monitoring [9].

Inter-Robot Interaction

For inter-robot interaction tasks, many participants used the shape of their hands to control the pattern formed with the robots which is also demonstrated in [34]. Contrary to the robot selection tasks, there was a significant increase ($p < 0.001$) in the use of two-handed gesture and a decrease in the use of one/two finger and one-handed gesture. Participants relied more on iconic and physical gestures and less on abstract, deictic, and symbolic gestures to control the robots' formations ($p < 0.001$). Similar to robot selection tasks, most gestures were discrete and robots-centric. These interactions can be used in applications like animation display [2] where it is critical to control the patterns formed by the robots.

Navigation

For navigation tasks many participants mapped movement of the robot(s) to hand motion, a similar trend as shown in [6, 17]. We also observed a significant increase ($p < 0.05$) in multi-modal interaction, specifically gesture combined with verbal commands. As shown in Figure 10, we saw significant

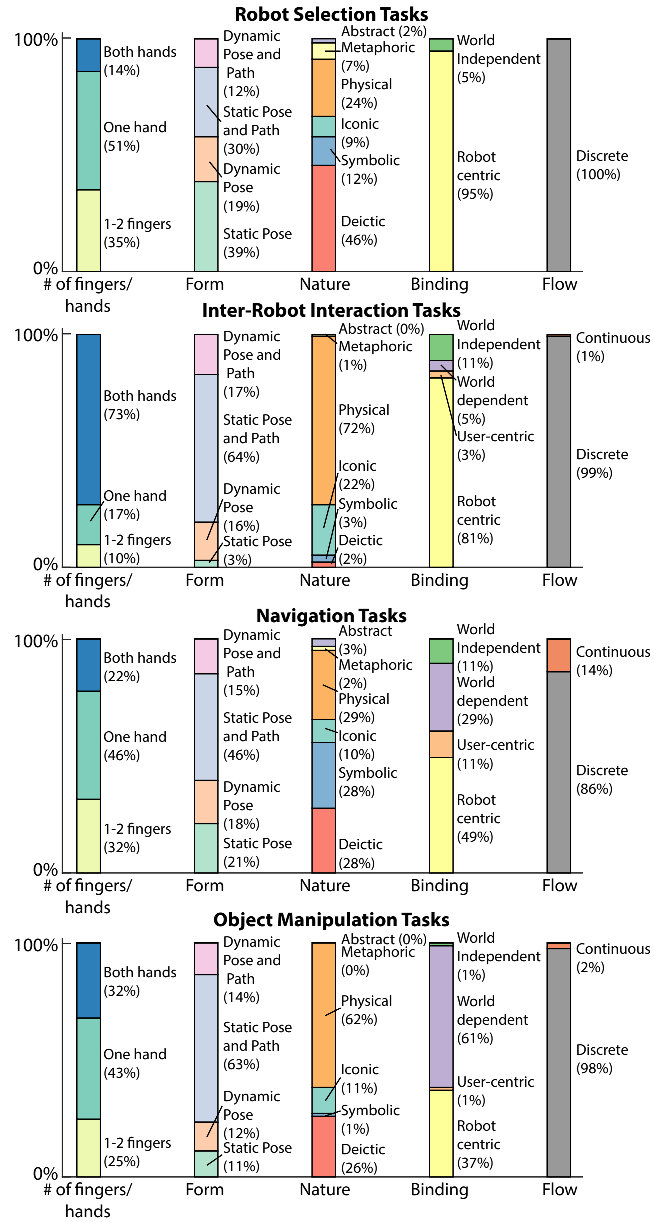

Figure 10. Taxonometric breakdown for different referent categories.

increases in one/two finger and one-handed gestures and a decrease in two-handed gesture ($p < 0.001$). Participants used more deictic and symbolic gestures and less physical gesture ($p < 0.001$). Presumably due to the nature of the tasks, there was a significant increase in continuous flow and a decrease in discrete flow ($p < 0.001$). These results can help inform the design of navigation control for deployment in search-and-rescue [44] and mining or agricultural foraging tasks [19].

Object Manipulation

Some participants explicitly communicated that they wanted the robots to push or grab the object through a tap or voice command, while others simply pushed or gestured the robots to move toward the object. There was a significant increase in touch + verbal interactions for the object manipulation tasks ($p < 0.05$). As may be expected as the tasks involved physical manipulation of an object, we saw increases in physical nature ($p < 0.001$) and world-dependent binding ($p < 0.001$). These results are relevant for a number of different applications such as the development of domestic robots for cleaning [33] or for robotic assembly tasks [54].

Design Insights

Based on the results of this study, this section presents a brief series of insights towards more effective interface design for future developers of swarm systems.

Our user-defined interaction set suggests that the interaction vocabulary changes depending on the state of the robots. Specifically, we observe that the number of robots as well as their proximity affects the user's interaction. This dynamic interaction vocabulary means that in addition to being able to detect input, swarm state information needs to be constantly relayed to the interface device and combined with an inference of user's intention in order to determine *contextually relevant* control inputs.

Gesture was the most commonly used interaction modality (66%) and this mirrors prior works in human-drone interaction [6, 1]. This suggests that if one were to choose only one type of sensor, one should choose a sensor that can detect different types of gestures. Interestingly, prior works in swarm robot control were able to correctly choose gesture as their main interaction modality even without such an elicitation study [15, 3]. Our study results better inform what types of gestures the sensor should be able to detect in order to better accommodate user's preference. For instance, being able to sense both hands is important when the user needs to control different number of robots as our results show a positive correlation between the number of robots and the number of fingers/hands used. Simultaneously, there is a need to detect relatively fine gestures (e.g., those involving only one or two fingers) as 31% of user interactions fell in that category.

While users heavily relied on gesture, they also used touch and verbal interactions 23% and 26% of the time. An ideal interface would be able to detect various types of touch and verbal interactions in addition to gesture. This would not only better support our user-defined interaction set but also provide users with additional degrees of freedom to leverage for different operational circumstances. For instance, in a dark room where the location of the robots is unknown, a user may find verbal interaction more useful than gesture or touch for getting attention of the robots.

Robots at the scale used in this study will struggle with the payload and energy demands of a vision system capable of user gesture identification, so even consensus-based approaches which take into account non-optimal classification may not be feasible. While centralized computer vision solutions (ideally incorporated into the infrastructure for robot path planning and control) may be the solution for tabletop and other stationary deployment environments, a gesture recognition device wearable by the operator may make the most sense for unstructured or mobile applications. Based on our finding that operators begin to use more two-handed gestures when robot number increases, a future wearable solution must be able to accommodate use of both arms/hands.

We found a negative correlation between increasing number of robots and proximity from the robots and self-reported ratings of interaction Simplicity and Precision. This finding aligns with prior research in teleoperated swarms, where users have a difficult time predicting how their control input will propagate through swarms [17]. Future interfaces should be "predictive" [49], providing some amount of feedback in real time to the user in the form of overlaid visual output from a path-planning algorithm or haptic feedback through the interface device, in order to decrease this uncertainty.

LIMITATIONS & FUTURE WORK

The fact that our study was conducted with relatively small tabletop robots limits the generalizability of our results. For example, the size of the robots discouraged several participants (P5, P7, P8, P15) from interacting with physical touch as they were "scared to break them" [P5, P8] even though they were told not to worry about damage to the robots. Limiting the robot environment to the tabletop also sets bounds on the maximum distance from the user as well as the maximum number of robots that can be interacted with at a time. Future work should investigate interaction in a room-scale environment – not only would it add more potential robots and distance, but also another dimension to vary (i.e., workspace or group height relative to user).

There exists a "legacy effect" in elicitation studies that leads users to fall back on their early or first responses even when parameters or tasks are varied in the future [28]. A larger participant pool would help to disambiguate user responses from this effect in future studies.

It is possible that the high percentage of users who elected for gestural control of the robots was influenced by the fact that our study was limited to tasks where the human operator is solely engaged with the robots. Future work could investigate whether preferred user input modality is changed if, for example, their visual attention is required elsewhere or hand(s) are otherwise occupied with some task.

Prior work has shown differences in preferred user input modality depending on the cultural background [11]. We did not specifically investigate this effect or account for it in our elicitation, although it is an important area for future work.

CONCLUSION

Mirroring the research that has shown a spectrum of feasible higher-level control strategies for swarm systems depending on their implementation details and level of autonomy, here we show that user-elicited interaction methods are closely related with the number of robots being interacted with at a time and their relative proximity. As future encountered robot swarms will be dynamic and mobile, our work indicates that their effective operation will also require dynamic, state-dependent interaction vocabularies to match them.

ACKNOWLEDGMENTS

We thank all the volunteers who participated in the study. This work was supported by Hasso Plattner Design Thinking Research Program, Samsung Scholarship for L.K., and Intelligence Community Postdoctoral Research Fellowship Program for D.D.

REFERENCES

[1] Parastoo Abtahi, David Y Zhao, Jane L. E, and James A Landay. 2017. Drone near me: Exploring touch-based human-drone interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 34.

[2] Javier Alonso-Mora, Andreas Breitenmoser, Martin Rufli, Roland Siegwart, and Paul Beardsley. 2012. Image and animation display with multiple mobile robots. *The International Journal of Robotics Research* 31, 6 (2012), 753–773.

[3] Javier Alonso-Mora, S Haegeli Lohaus, Philipp Leemann, Roland Siegwart, and Paul Beardsley. 2015. Gesture based human-multi-robot swarm interaction and its application to an interactive display. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5948–5953.

[4] Sean Braley, Calvin Rubens, Timothy R Merritt, and Roel Vertegaal. 2018. GridDrones: A self-levitating physical voxel lattice for 3D surface deformations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, D200.

[5] Ian M Bullock and Aaron M Dollar. 2011. Classifying human manipulation behavior. In *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE, 1–6.

[6] Jessica R Cauchard, Jane L. E, Kevin Y Zhai, and James A Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 361–365.

[7] Luca Consolini, Fabio Morbidi, Domenico Prattichizzo, and Mario Tosques. 2008. Leader–follower formation control of nonholonomic mobile robots with input constraints. *Automatica* 44, 5 (2008), 1343–1349.

[8] Gilles Coppin and François Legras. 2011. Autonomy spectrum and performance perception issues in swarm supervisory control. *Proc. IEEE* 100, 3 (2011), 590–603.

[9] Miguel Duarte, Jorge Gomes, Vasco Costa, Tiago Rodrigues, Fernando Silva, Víctor Lobo, Mario Monteiro Marques, Sancho Moura Oliveira, and Anders Lyhne Christensen. 2016. Application of swarm robotics systems to marine environmental monitoring. In *OCEANS 2016-Shanghai*. IEEE, 1–8.

[10] Gregory Dudek, Michael Jenkin, and Evangelos Milios. 2002. A taxonomy of multirobot systems. *Robot teams: From diversity to polymorphism* (2002), 3–22.

[11] Jane L E, Ilene L E, James A Landay, and Jessica R Cauchard. 2017. Drone & wo: Cultural influences on human-drone interaction techniques. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6794–6799.

[12] George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais. 1987. The vocabulary problem in human-system communication. *Commun. ACM* 30, 11 (1987), 964–971.

[13] Antonio Gomes, Calvin Rubens, Sean Braley, and Roel Vertegaal. 2016. Bitdrones: Towards using 3d nanocopter displays as interactive self-levitating programmable matter. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 770–780.

[14] Michael D Good, John A Whiteside, Dennis R Wixon, and Sandra J Jones. 1984. Building a user-derived interface. *Commun. ACM* 27, 10 (1984), 1032–1043.

[15] Boris Gromov, Luca M Gambardella, and Gianni A Di Caro. 2016. Wearable multi-modal interface for human multi-robot interaction. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 240–245.

[16] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 234–241.

[17] Geraint Jones, Nadia Berthouze, Roman Bielski, and Simon Julier. 2010. Towards a situated, multimodal interface for multiple UAV control. In *2010 IEEE International Conference on Robotics and Automation*. IEEE, 1739–1744.

[18] Jun Kato, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2009. Multi-touch interface for controlling multiple mobile robots. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 3443–3448.

[19] Belkacem Khaldi and Foudil Cherif. 2015. An overview of swarm robotics: Swarm intelligence applied to multi-robotics. *International Journal of Computer Applications* 126, 2 (2015).

[20] Lawrence H Kim and Sean Follmer. 2017. Ubiswarm: Ubiquitous robotic interfaces and investigation of abstract motion as a display. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 66.

[21] Lawrence H Kim and Sean Follmer. 2019. SwarmHaptics: Haptic Display with Swarm Robots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 688.

[22] Andreas Kolling, Katia Sycara, Steven Nunnally, and Michael Lewis. 2013. Human-swarm interaction: An experimental study of two types of interaction with foraging swarms. *Journal of Human-Robot Interaction* 2, 2 (2013), 103–129.

[23] Andreas Kolling, Phillip Walker, Nilanjan Chakraborty, Katia Sycara, and Michael Lewis. 2015. Human interaction with robot swarms: A survey. *IEEE Transactions on Human-Machine Systems* 46, 1 (2015), 9–26.

[24] Mathieu Le Goc, Lawrence H Kim, Ali Parsaei, Jean-Daniel Fekete, Pierre Dragicevic, and Sean Follmer. 2016. Zooids: Building blocks for swarm user interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 97–109.

[25] Mathieu Le Goc, Charles Perin, Sean Follmer, Jean-Daniel Fekete, and Pierre Dragicevic. 2018. Dynamic composite data physicalization using wheeled micro-robots. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 737–747.

[26] John H McDonald. 2009. *Handbook of biological statistics*. Vol. 2. sparky house publishing Baltimore, MD.

[27] James McLurkin, Jennifer Smith, James Frankel, David Sotkowitz, David Blau, and Brian Schmidt. 2006. Speaking Swarmish: Human-Robot Interface Design for Large Swarms of Autonomous Mobile Robots.. In *AAAI spring symposium: to boldly go where no human-robot team has gone before*, Vol. 72.

[28] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, Jacob O Wobbrock, and others. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.

[29] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*. Canadian Information Processing Society, 261–268.

[30] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 331–338.

[31] Jawad Nagi, Alessandro Giusti, Luca M Gambardella, and Gianni A Di Caro. 2014. Human-swarm interaction using spatial gestures. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3834–3841.

[32] Ayberk Özgür, Séverin Lemaignan, Wafa Johal, Maria Beltran, Manon Briod, Léa Pereyre, Francesco Mondada, and Pierre Dillenbourg. 2017. Cellulo: Versatile handheld robots for education. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 119–127.

[33] Jordi Palacin, José Antonio Salse, Ignasi Valgañón, and Xavi Clua. 2004. Building a mobile robot for a floor-cleaning operation in domestic environments. *IEEE Transactions on instrumentation and measurement* 53, 5 (2004), 1418–1424.

[34] Ekaterina Peshkova and Martin Hitz. 2017. Exploring user-defined gestures to control a group of four UAVs. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 169–174.

[35] Vagelis Plevris, Matthew G Karlaftis, and Nikos D Lagaros. 2010. A swarm intelligence approach for emergency infrastructure inspection scheduling. In *Sustainable and Resilient Critical Infrastructure Systems*. Springer, 201–230.

[36] Gaëtan Podevijn, Rehan O'grady, Nithin Mathews, Audrey Gilles, Carole Fantini-Hauwel, and Marco Dorigo. 2016. Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction. *Swarm Intelligence* 10, 3 (2016), 193–210.

[37] Gaëtan Podevijn, Rehan O'Grady, Youssef SG Nashed, and Marco Dorigo. 2013. Gesturing at subswarms: Towards direct human control of robot swarms. In *Conference Towards Autonomous Robotic Systems*. Springer, 390–403.

[38] Erol Şahin. 2004. Swarm robotics: From sources of inspiration to domains of application. In *International workshop on swarm robotics*. Springer, 10–20.

[39] John R Searle. 1975. A taxonomy of illocutionary acts. (1975).

[40] Shiva Shahrokhi and Aaron T Becker. 2016. Object manipulation and position control using a swarm with global inputs. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 561–566.

[41] Thomas B Sheridan and William L Verplank. 1978. *Human and computer control of undersea teleoperators*. Technical Report. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

[42] William M Spears, Diana F Spears, Jerry C Hamann, and Rodney Heil. 2004. Distributed, physics-based control of swarms of vehicles. *Autonomous Robots* 17, 2-3 (2004), 137–162.

[43] Adrian Stoica, Theodoros Theodoridis, Huosheng Hu, Klaus McDonald-Maier, and David F Barrero. 2013. Towards human-friendly efficient control of multi-robot teams. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 226–231.

[44] Daniel P Stormont. 2005. Autonomous rescue robot swarms for first responders. In *CIHSPS 2005. Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2005.* IEEE, 151–157.

[45] Ryo Suzuki, Jun Kato, Mark D Gross, and Tom Yeh. 2018. Reactile: Programming Swarm User Interfaces through Direct Physical Manipulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 199.

[46] Ryo Suzuki, Clement Zheng, Yasuaki Kakehi, Tom Yeh, Ellen Yi-Luen Do, Mark D Gross, and Daniel Leithinger. 2019. ShapeBots: Shape-changing Swarm Robots. *arXiv preprint arXiv:1909.03372* (2019).

[47] Leila Takayama and Caroline Pantofaru. 2009. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5495–5502.

[48] J Triesch. 1998. Ch. Von der Malsburg. A Gesture Interface for Human-Robot Interaction. In *Proc. of the 2nd Int. Conf. on Automatic Face and Gesture Recognition*. 546–551.

[49] Phillip Walker, Steven Nunnally, Mike Lewis, Andreas Kolling, Nilanjan Chakraborty, and Katia Sycara. 2012. Neglect benevolence in human control of swarms in the presence of latency. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3009–3014.

[50] Mark Weiser. 1993. Ubiquitous computing. *Computer* 10 (1993), 71–72.

[51] Andrew D Wilson, Shahram Izadi, Otmar Hilliges, Armando Garcia-Mendoza, and David Kirk. 2008. Bringing physics to the surface. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. ACM, 67–76.

[52] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*. ACM, 1869–1872.

[53] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1083–1092.

[54] Yiwei Zhao, Lawrence H Kim, Ye Wang, Mathieu Le Goc, and Sean Follmer. 2017. Robotic assembly of haptic proxy objects for tangible interaction and virtual reality. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ACM, 82–91.